

Query Match 100.0%; Score 2627.6; DB 6; Length 2628;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTGTCCTAGTGGGAAAGTCTTCCAAATGGGAGCTGGCCAGTGTGCACAAA 60
 Db 1 ATGGCCCTGTCCTAGTGGGAAAGTCTTCCAAATGGGAGCTGGCCAGTGTGCACAAA 60
 QY 61 AGCCTGACTGGGCAACAAGTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 Db 61 AGCCTGACTGGGCAACAAGTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 QY 121 CT 180
 Db 121 CT 180
 QY 181 TGTGTGCTGT 240
 Db 181 TGTGTGCTGT 240
 QY 241 TGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 Db 241 TGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 QY 301 ATTTGCTGT 360
 Db 301 ATTTGCTGT 360
 QY 361 CCCAGTGTGCTGT 420
 Db 361 CCCAGTGTGCTGT 420
 QY 421 ATCTTCTGT 480
 Db 421 ATCTTCTGT 480
 QY 481 GACCGCTGT 540
 Db 481 GACCGCTGT 540
 QY 541 CCAGATCCCCGGGAGGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 Db 541 CCAGATCCCCGGGAGGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 QY 601 AAGGGAAGATCTTTGACAGAGTGGGCACTTCTGTGTGTGTGTGTGTGTGTGTGTGT 660
 Db 601 AAGGGAAGATCTTTGACAGAGTGGGCACTTCTGTGTGTGTGTGTGTGTGTGTGTGT 660
 QY 661 CGGACCCACTCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 Db 661 CGGACCCACTCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 QY 721 GTTGGCCACCGTGGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
 Db 721 GTTGGCCACCGTGGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
 QY 781 GGGCGGTGCTGT 840
 Db 781 GGGCGGTGCTGT 840
 QY 841 TACAGGGGTCTGT 900
 Db 841 TACAGGGGTCTGT 900
 QY 901 CCTGGGAGCTGT 960
 Db 901 CCTGGGAGCTGT 960
 QY 961 ATGGCCGCTGT 1020
 Db 961 ATGGCCGCTGT 1020

QY 1021 TCGGAGAGGTCCTGCTGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 Db 1021 TCGGAGAGGTCCTGCTGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 QY 1081 GCGGTCGT 1140
 Db 1081 GCGGTCGT 1140
 QY 1141 CTGAAGGCTGT 1200
 Db 1141 CTGAAGGCTGT 1200
 QY 1201 CAGCTAGAGATGT 1260
 Db 1201 CAGCTAGAGATGT 1260
 QY 1261 CTGT 1320
 Db 1261 CTGT 1320
 QY 1321 GACATCTGT 1380
 Db 1321 GACATCTGT 1380
 QY 1381 GGGCCCTATGT 1440
 Db 1381 GGGCCCTATGT 1440
 QY 1441 GCACAGCCAGGAGGCTTTTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
 Db 1441 GCACAGCCAGGAGGCTTTTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
 QY 1501 GCGGAGCCACCCCTCATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
 Db 1501 GCGGAGCCACCCCTCATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
 QY 1561 GTCAAGACCCACTGT 1620
 Db 1561 GTCAAGACCCACTGT 1620
 QY 1621 GATGGACCTACATGT 1680
 Db 1621 GATGGACCTACATGT 1680
 QY 1681 GCTGTCCAGGAGGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740
 Db 1681 GCTGTCCAGGAGGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740
 QY 1741 TTTAAGGTGCAACCTATACCGGAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
 Db 1741 TTTAAGGTGCAACCTATACCGGAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
 QY 1801 GTGGGCTGT 1860
 Db 1801 GTGGGCTGT 1860
 QY 1861 GCTGT 1920
 Db 1861 GCTGT 1920
 QY 1921 ACCGACCTACGCGCCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
 Db 1921 ACCGACCTACGCGCCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
 QY 1981 AACCTGTGCTGT 2040
 Db 1981 AACCTGTGCTGT 2040
 QY 2041 GCCACCATGT 2100
 Db 2041 GCCACCATGT 2100
 QY 2101 CCCAGCTGT 2160

[illegible]

Db	675	CTCCCTCTCTCCCTCCCTCTTCCCGCCAGAGATGGGGTGGACCACTGCTTATAG	734
Qy	181	TGTGAGCGCTGTGTGCACAGACAGAACTACACCCGGCGCAGCGGGAGCAAGAGGGCGGC	240
Db	735	TGTGTGCGCTGTGTGCACAGACAGAACTACACCCGGCGCGACGGGAGCAAGGGCGGC	794
Qy	241	TCTGAGGACAGCGCGAGCGGGTGGCCCTCTCACTTCCCGAGATGGCCCGCCCTTGCGCC	300
Db	795	TCTGAGGACAGCGCGAGCGGGTGGCCCTCTCACTTCCCGAGATGGCCCGCCCTTGCGCC	854
Qy	301	ATTTCCTGGCTGTAGAGGGGCTCTCTGCTCATATGGGGTATGCTTGAGCCCGACCCGCGC	360
Db	855	ATTTCCTGGCTGTATAGGGGGCTCTGCTCATATGGGGTATGCTTGAGCCCGACCCGCGC	914
Qy	361	CCAGATGGCCCGGCTCGGCTCTCACTTCCAGAGAACCTCTGTCTGCCAACGCTCTGCGC	420
Db	915	CCAGATGGCCCGGCTCGGCTCTCACTTCCAGAGAACCTCTGTCTGCCAACGCTCTGCGC	974
Qy	421	ATCTTCTGGGCCCCCAGGGCTCCTTGAACCTCCAGGCGCATGTACTAGATGATTACGA	480
Db	975	ATCTTCTGGGCCCCCAGGGCTCCTTGAACCTCCAGGCGCATGTACTAGATGATTACGA	1034
Qy	481	GACCGGCTCTTCTGGGTGGCTTGAGACGCGCTCTACTCTCTGCGGCTTGAGACAGGCAATGG	540
Db	1035	GACCGGCTCTTCTGGGTGGCTTGAGACGCGCTCTACTCTCTGCGGCTTGAGACAGGCAATGG	1094
Qy	541	CCAGATCCCCGGAGGTCTCTGTGTGACCCGCGAGCGGAGAGAGAGAGTGTTCGA	600
Db	1095	CCAGATCCCCGGAGGTCTCTGTGTGACCCGCGAGCGGAGAGAGAGTGTTCGA	1154
Qy	601	AAGGAGAGATCTTGTGACAGATCGGCCAATTCGTGCGGGGTCTACAGCTTCACAC	660
Db	1155	AAGGAGAGATCTTGTGACAGATCGGCCAATTCGTGCGGGGTCTACAGCTTCACAC	1214
Qy	661	CGGACCACTCTAGCTTGTGGTACCTGGGGGCTTCAGGCCACCTGATCCCTCATCA	720
Db	1215	CGGACCACTCTAGCTTGTGGTACCTGGGGGCTTCAGGCCACCTGATCCCTCATCA	1274
Qy	721	GTTGGCCACGCTGGGGAGCATGTGCTTCACCTGAGCGTGTGGAAAGTGGCGG	780
Db	1275	GTTGGCCACGCTGGGGAGCATGTGCTTCACCTGAGCGTGTGGAAAGTGGCGG	1334
Qy	781	GGGCGGTGCTCTACAGAGCCGAGCGTCCCTTGGCCAGCACTTACAGCGGGAGCTG	840
Db	1335	GGGCGGTGCTCTACAGAGCCGAGCGTCCCTTGGCCAGCACTTACAGCGGGAGCTG	1394
Qy	841	TACAGGGGTCTACTGTACTTCTCTGGGGCGAAGGCGCATGTATCTTCCGAATGGAGGT	900
Db	1395	TACAGGGGTCTACTGTACTTCTCTGGGGCGAAGGCGCATGTATCTTCCGAATGGAGGT	1454
Qy	901	CCTGGGCGAGCTCTGCTTCCGACTTACACAGAGTCTTGTGCAGCAAGCCCGGTTTGTG	960
Db	1455	CCTGGGCGAGCTCTGCTTCCGACTTACACAGAGTCTTGTGCAGCAAGCCCGGTTTGTG	1514
Qy	961	ATGGCCGCCGAGTCCCTGTAGAATCTGTGACCAAGACATGTGCAAGGGTACTTCTTCTTC	1020
Db	1515	ATGGCCGCCGAGTCCCTGTAGAATCTGTGACCAAGACATGTGCAAGGGTACTTCTTCTTC	1574
Qy	1021	TGGGAGACGCTCCCTGCGGCCATGTGAGTGTGCAATGTCTACTGTACGCGCGTGGGC	1080
Db	1575	TGGGAGACGCTCCCTGCGGCCATGTGAGTGTGCAATGTCTACTGTACGCGCGTGGGC	1634
Qy	1081	CGGCTCTGCTGATGATGTCTGTGGGGCGAGGGGGTGTGGTAAATAATGAGCACTTTC	1140
Db	1635	CGGCTCTGCTGATGATGTCTGTGGGGCGAGGGGGTGTGGTAAATAATGAGCACTTTC	1694
Qy	1141	CTCAAGGCGAGGCGTGTCTGTGTGCGCGGCGCTGTGTGTGCCGAGACCACTTGTAC	1200
Db	1695	CTCAAGGCGAGGCGTGTCTGTGTGCGCGGCGCTGTGTGTGCCGAGACCACTTGTAC	1754
Qy	1201	CAGCTAAGATGTGTCTGTGTGTGTGGCCCAAGGGCGGGAAGAGCTCTCAGGTGTACCG	1260

QY	2341	AAGCCAGAGAGACCCCCCAAGCCCGGGAGAGCTGGCTTTCACCCCAACCCAAAGGCTGTGTAC	2400
Db	2895	AAGCCAGAGAGACCCCCCAAGCCCGGGAGAGCTGGCTTTCACCCCAACCCAAAGGCTGTGTAC	2954
QY	2401	AAGGACATCTGTGACGCTATTGGCTGGCCCAACCTGGCCGGGGGATGATGATCTGTGTAC	2460
Db	2955	AAGGACATCTGTGACGCTATTGGCTGGCCCAACCTGGCCGGGGGATGATGATCTGTGTAC	3014
QY	2461	CGCGTGTGTACAGGGGCAACCAAGCAATGCTCAGAGCTGTTCGGAGCCGAGCCGGGGC	2520
Db	3015	CGCGTGTGTGTCAAGGGGCACCCAGGGAATGCTCAGAGCTGTTCGGAGCCGAGCCGGGGC	3074
QY	2521	AAGCAGGCGCAAGGCGCAAGAGCTGGGCAAGGGCTGTAGCTAGGCAAGAAATGACAGAGCCGG	2580
Db	3075	AAGCAGGCGCAAGGCGCAAGAGCTGGGCAAGGGCTGTAGCTAGGCAAGAAATGACAGAGCCGG	3134
QY	2581	GTGCATGCGCAGACAAATCGAGCCCGCCGGGAGGTGGAGGCGCCAGTAG	2628
Db	3135	GTGCATGCGCAGACAAATCGAGCCCGCCGGGAGGTGGAGGCGCCAGTAG	3182

RESULT 3	AX253547	2349 bp	DNA	linear	PAT 10-OCT-2001
LOCUS	AX253547				
DEFINITION	Sequence 3 from Patent WO0170806.				
ACCESSION	AX253547				
VERSION	AX253547.1	GI:16074049			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2349)				
AUTHORS	Walke,D.W., Wilganowski,N.L., Turner,C.A., Hilbun,E., Wang,X., Donoho,G. and Scoville,J.				
TITLE	Human secreted proteins and polynucleotides encoding the same Patent: WO 0170806-A 3 27-SEP-2001;				
JOURNAL	Lexicon Genetics Incorporated (US)				
FEATURES	Location/Qualifiers				
source	1..2349				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	406 a 747 c 753 g 442 t 1 others				
ORIGIN					
Query Match	89.4%; Score 2348.6; DB 6; Length 2349;				
Best Local Similarity	100.0%; Pred. NO. 0;				
Matches 2349;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	280 ATGGCCCCCTGGCCCTGGGCCATTGCTGCTGCTAGGGGCTCCTGTCATGAGGGGT 339				
1	ATGGCCCCCTGGCCCTGGGCCATTGCTGCTGCTGCTAGGGGCTCCTGTCATGAGGGGT 60				
Qy	340 AGCTCTGGCCCCCAGCCCCCGCCAGTGTGGCCCGCCGCTGGGGCTCTCTACGACATC 399				
61	AGCTCTGGCCCCCAGCCCCCGCCAGTGTGGCCCGCCGCTGGGGCTCTCTACGACATC 120				
Qy	400 CTGCTGCGAACCGCTCTGCCATCTTTC7GTGGCCCCCAGGGGCTCCCTGAACCTCCAGGCC 459				
121	CTGCTGCGAACCGCTCTGCCATCTTTC7GTGGCCCCCAGGGGCTCCCTGAACCTCCAGGCC 180				
Qy	460 ATGTACTAGATGATGATACCGAGACCGCCCTTTC7GTGGGGGCGCTGGAGCGCCCTTACTCT 519				
Db	181 ATGTACTAGATGATGATACCGAGACCGCCCTTTC7GTGGGGGCGCTGGAGCGCCCTTACTCT 240				
Qy	520 CTGGGGCTGAGCCAGGCGATGGCCAGATCCCGGGAGGTCCCTGTGGCCACCGAGCCAGGA 579				
Db	241 CTGGGGCTGAGCCAGGCGATGGCCAGATCCCGGGAGGTCCCTGTGGCCACCGAGCCAGGA 300				
Qy	580 CAGAGGAGAGATGTGTTCGAAGGGAAGATCCTTGACAGAGTGGCCCAACTTCTGTG 639				
Db	301 CAGAGGAGAGATGTGTTCGAAGGGAAGATCCTTGACAGAGTGGCCCAACTTCTGTG 360				

QY 640 CGGGTCTACAGCTTCACACCGGACCCACCTGTTAGCTGTGGAGCTGGGGCTTCCAG 699
 Db 361 CGGGTCTACAGCTTCACACCGGACCCACCTGTTAGCTGTGGAGCTGGGGCTTCCAG 420
 QY 700 CCCACTGTGGCCCTCATCAGTTGGCCACCGTGGGGAGCATGTCTCCACCTGGAGCT 759
 Db 421 CCCACTGTGGCCCTCATCAGTTGGCCACCGTGGGGAGCATGTCTCCACCTGGAGCT 480
 QY 760 GGCAGTGTGAAAGTGGCGGGGGGGCTGCTCCACAGAGCCAGCCGCTTCCAGC 819
 Db 481 GGCAGTGTGAAAGTGGCGGGGGGGCTGCTCCACAGAGCCAGCCGCTTCCAGC 540
 QY 820 ACCTTCATAGAGGGGGAGCTGTACAGAGGCTCTACCTGTGACTTCTGGGGGAGAGGCC 879
 Db 541 ACCTTCATAGAGGGGGAGCTGTACAGAGGCTCTACCTGTGACTTCTGGGGGAGAGGCC 600
 QY 880 ATGATCTTCCGAAAGTGAAGTCTCTGGCCAGCTCTGAGTCCGAGCTCTACAGAGCTCTC 939
 Db 601 ATGATCTTCCGAAAGTGAAGTCTCTGGCCAGCTCTGAGTCCGAGCTCTACAGAGCTCTC 660
 QY 940 TTGACAGACCCCGGTTGTGATGGCCCGGATCCCTGAGAACTCTGACACAGAGCAAT 999
 Db 661 TTGACAGACCCCGGTTGTGATGGCCCGGATCCCTGAGAACTCTGACACAGAGCAAT 720
 QY 1000 GACAAAGTGTACTTCTTCTCTGAGAGAGGTTCCCTGGCCGAGATGAGTGGCTCGAACCAT 1059
 Db 721 GACAAAGTGTACTTCTTCTCTGAGAGAGGTTCCCTGGCCGAGATGAGTGGCTCGAACCAT 780
 QY 1060 GTACAGTGAAGCGGGTGGGGCGGCTGCTGAGTGAATGATGCTGGGGGCGACAGGGGTCTG 1119
 Db 781 GTACAGTGAAGCGGGTGGGGCGGCTGCTGAGTGAATGATGCTGGGGGCGACAGGGGTCTG 840
 QY 1120 GTGAAACAATGAGACACTTCTCTCAAGAGCCAGGAGTGTCTCGTCCGCTCCGCTCGT 1179
 Db 841 GTGAAACAATGAGACACTTCTCTCAAGAGCCAGGAGTGTCTCGTCCGCTCCGCTCGT 900
 QY 1180 GGTGCGAGAGCCACTTGTGACAGCTAGAGATGTCTCTGTGTGGCCCAAGGCGGG 1239
 Db 901 GGTGCGAGAGCCACTTGTGACAGCTAGAGATGTCTCTGTGTGGCCCAAGGCGGG 960
 QY 1240 AAGAGCTCGAGGTGTACGGGCTTTCAGACAGCTGAGTCCGTTCAGAGGCTTCCGCT 1299
 Db 961 AAGAGCTCGAGGTGTACGGGCTTTCAGACAGCTGAGTCCGTTCAGAGGCTTCCGCT 1020
 QY 1300 GTCTGTGTGTACCAATGAGAGCATCTGGAGGTTTCAAGGGCCCTTGGCCACCGA 1359
 Db 1021 GTCTGTGTGTACCAATGAGAGCATCTGGAGGTTTCAAGGGCCCTTGGCCACCGA 1080
 QY 1360 GATGGGCTCAGACCACTGAGGGGCGCTATGAGGGGCAAGGTGCTTCCGCTCGCTGGC 1419
 Db 1081 GATGGGCTCAGACCACTGAGGGGCGCTATGAGGGGCAAGGTGCTTCCGCTCGCTGGC 1140
 QY 1420 GTGTGCGCCAGAGATGAGCCAGAGCCAGAGCGGCTTTTGGGAGCAGCAACAGACTAC 1479
 Db 1141 GTGTGCGCCAGAGATGAGCCAGAGCCAGAGCGGCTTTTGGGAGCAGCAACAGACTAC 1200
 QY 1480 CCAGATGAGTGTGAGTTCGCCGAGGCGCCCTCATGTTTGTGGCTGTGGGCT 1539
 Db 1201 CCAGATGAGTGTGAGTTCGCCGAGGCGCCCTCATGTTTGTGGCTGTGGGCT 1260
 QY 1540 CGAGATGAGGCGGCTGTCTTGTCAAGAGCCACCTGAGCCAGACTACACAGATGCTG 1599
 Db 1261 CGAGATGAGGCGGCTGTCTTGTCAAGAGCCACCTGAGCCAGACTACACAGATGCTG 1320
 QY 1600 GTGAGCGCGTGAAGCAGAGAGTGGAGCTTACATGTCATTTCTTGGGAGTACTCA 1659
 Db 1321 GTGAGCGCGTGAAGCAGAGAGTGGAGCTTACATGTCATTTCTTGGGAGTACTCA 1380
 QY 1660 GGGTGTGTGTCAAGTCAATGCTCTCAGAGGAGGGGCTCAGCTGAACCTGAGAAATG 1719
 Db 1381 GGGTGTGTGTCAAGTCAATGCTCTCAGAGGAGGGGCTCAGCTGAACCTGAGAAATG 1440

QY 1720 GTTCTGAGAGAGCTCCAGGTGTTTAAAGTGGCAACCTATACCCGAATGAGATCTCT 1779
 Db 1441 GTTCTGAGAGAGCTCCAGGTGTTTAAAGTGGCAACCTATACCCGAATGAGATCTCT 1500
 QY 1780 GTCAAAAGCAAAATGCTATACGTGGGCTCTCGGCTGTGGTGGCCAGAGCTGGCTGAC 1839
 Db 1501 GTCAAAAGCAAAATGCTATACGTGGGCTCTCGGCTGTGGTGGCCAGAGCTGGCTGAC 1560
 QY 1840 CAATGTAGACTTACAGCACTGCTGTGAGAGTGTCTGGGGCCGGAGCCCAATACGT 1899
 Db 1561 CAATGTAGACTTACAGCACTGCTGTGAGAGTGTCTGGGGCCGGAGCCCAATACGT 1620
 QY 1900 GCTTGGATGTGCTCTCTCTTACCACTACCGCCAGCTTGGCAAGCGGCTTCCGC 1959
 Db 1621 GCTTGGATGTGCTCTCTCTTACCACTACCGCCAGCTTGGCAAGCGGCTTCCGC 1680
 QY 1960 CGGAGAGCATCCGGCAGCGCAACCTGCTGAGTGGCTGGGCGAGAGCCAGAGAA 2019
 Db 1681 CGGAGAGCATCCGGCAGCGCAACCTGCTGAGTGGCTGGGCGAGAGCCAGAGAA 1740
 QY 2020 GAGGAGTGGGAGTGTGGGAGCCAGCATGCTTACGGCAGAGCAATAGCACTTTC 2079
 Db 1741 GAGGAGTGGGAGTGTGGGAGCCAGCATGCTTACGGCAGAGCAATAGCACTTTC 1800
 QY 2080 CTGAGTGTGCTCCCACTCTCCCACTGCTGTGGGCTGTGGGCTTGGAGCCAGGG 2139
 Db 1801 CTGAGTGTGCTCCCACTCTCCCACTGCTGTGGGCTGTGGGCTTGGAGCCAGGG 1860
 QY 2140 GATGAGGGGCTTACAGAGTGAAGAGAGAGAGAGAGAGTGTGACACAGAGCGGGGCTG 2199
 Db 1861 GATGAGGGGCTTACAGAGTGAAGAGAGAGAGAGAGTGTGACACAGAGCGGGGCTG 1920
 QY 2200 CTGTTCCGAGGCTTACCGCTTTCATGCGGGGACCTTACCTGACCTCTGAGCAT 2259
 Db 1921 CTGTTCCGAGGCTTACCGCTTTCATGCGGGGACCTTACCTGACCTCTGAGCAT 1980
 QY 2260 GGGTGTCCAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2319
 Db 1981 GGGTGTCCAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
 QY 2320 AACCTGTCTCCCGAGAGCAAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2379
 Db 2041 AACCTGTCTCCCGAGAGCAAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 QY 2380 ACCGACCCAGAGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2439
 Db 2101 ACCGACCCAGAGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2440 CGGCTGATGATGCTGTGAGCGGTGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAG 2499
 Db 2161 CGGCTGATGATGCTGTGAGCGGTGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 QY 2500 TTCCGGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2559
 Db 2221 TTCCGGAGCGGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
 QY 2560 GGCAGAGATGATGAG 2619
 Db 2281 GGCAGAGATGATGAG 2340
 QY 2620 GCCAGCTAG 2628
 Db 2341 GCCAGCTAG 2349

RESULT 4
 AB029496 4700 bp mRNA linear PRI 07-JUL-2000
 LOCUS
 DEFINITION Homo sapiens mRNA for semaphorin sem2, complete cds.
 ACCESSION AB029496.1 GI:8978201
 VERSION semaphorin sem2.
 KEYWORDS semaphorin sem2, complete cds.
 SOURCE Homo sapiens cDNA to mRNA.

Db	1441	GTCTGGAGGAGCTCCAGAGTGTAAAGTGTCCACACACTTATCACCCGAATGAGATCTT	1500
OY	1780	GTCAAAAGGCAAAATGCTATACGTGGGCTCTCGGCTGGGTGTGGCCACAGCTGCGCTCCAC	1839
Db	1501	GTCAAAAGGCAAAATGCTTACGTGGGCTCTCGGCTGGGTGTGGCCACAGCTGCGCTCCAC	1560
OY	1840	CAATGTGAGACTTACGCGACCTGCTGTGCAAGTGTGCTGGGCCCGGGACCCCTACTGT	1899
Db	1561	CAATGTGAGACTTACGCGACCTGCTGTGCAAGTGTGCTGGGCCCGGGACCCCTACTGT	1620
OY	1900	GGCTGGAGTGTGGCTCTCGTACCCTACCGCCACCGCTTGGCAACGCGGTTCCGC	1959
Db	1621	GGCTGGAGTGTGGCTCTCGTACCCTACCGCCACCGCTTGGCAAGCGCGGTTCCGC	1680
OY	1960	CGGACGAGACATCCGGACCGGCAACCTGGCTCTGACAGTGGCTGGGCGAGAGCCAGAGAA	2019
Db	1681	CGGACGAGACATCCGGACCGGCAACCTGGCTCTGACAGTGGCTGGGCGAGAGCCAGAGAA	1740
OY	2020	GAGCAGTGGGACTTGTGGCAGCCACCATGTCTACGGACGAGAGACAAATGACCTTC	2079
Db	1741	GAGCAGTGGGACTTGTGGCAGCCACCATGTCTACGGACGAGAGACAAATGACCTTC	1800
OY	2080	CTGAGAGTCTGCGCCAAAGTCTCCCAAGCTGCTGTGGCGCTGGCTCTTGACAGAGCCAGG	2139
Db	1801	CTGAGAGTCTGCGCCAAAGTCTCCCAAGCTGCTGTGGCGCTGGCTCTTGACAGAGCCAGG	1860
OY	2140	GATGAGGGGCTGACACAGTGAAGACGAGACGAGAGATCTTGACACAGAGGAGGCGCTG	2199
Db	1861	GATGAGGGGCTGACACAGTGAAGACGAGACGAGAGATCTTGACACAGAGGAGGCGCTG	1920
OY	2200	CTGTTCCCGCAGGCTTAGCGCTTTCGATCGGGACCTACACTGCACACTCTGTAAGCAT	2259
Db	1921	CTGTTCCCGCAGGCTTAGCGCTTTCGATCGGGACCTACACTGCACACTCTGTAAGCAT	1980
OY	2260	GCGTTCTCCACAGCTGTGGTCCGCTGGCTCTGGGTGATGTGGCTACAGCTGTGAC	2319
Db	1981	GCGTTCTCCACAGCTGTGGTCCGCTGGCTCTGGGTGATGTGGCTACAGCTGTGAC	2040
OY	2320	AACCTTTCCCTCCGAGCCAAAGCCAGAGAGCCCCACAGCCCGGGGAGGCGCTGGCTTC	2379
Db	2041	AACCTTTCCCTCCGAGCCAAAGCCAGAGAGCCCCACAGCCCGGGGAGGCGCTGGCTTC	2100
OY	2380	ACCCACCCACAGGCTGTGTACAAAGACATCTGACACTATTGGCTTGGCCAACTGTGCC	2439
Db	2101	ACCCACCCACAGGCTGTGTACAAAGACATCTGACACTATTGGCTTGGCCAACTGTGCC	2160
OY	2440	CGGGTGATGACTGTGTGAGCGCTGTGTGTGAGGGGACCAACGGAAATGCTAGAGCTGC	2499
Db	2161	CGGGTGATGACTGTGTGAGCGCTGTGTGTGAGGGGACCAACGGAAATGCTAGAGCTGC	2220
OY	2500	TTCGGGACCGGAGCCGGGGCAACGAGCCACAGGGGCAAGAGCTGGCGAGGCTGTGAGCTA	2559
Db	2221	TTCGGGACCGGAGCCGGGGCAACGAGCCACAGGGGCAAGAGCTGGCGAGGCTGTGAGCTA	2280
OY	2560	GCGAAGAGATGAAGAGCCGGGTGCTATCCGAGACATCGAGAGCGCCCGGAGGTGTG	2619
Db	2281	GCGAAGAGATGAAGAGCCGGGTGCTATCCGAGACATCGAGAGCGCCCGGAGGTGTG	2340
OY	2620	GCCAGTAG 2628	
Db	2341	GCCAGTAG 2349	

RESULT 5			
AK024425	AK024425	4469 bp	linear
LOCUS			PRI 29-SEP-2000
DEFINITION	Homo sapiens mRNA for FLJ00014 protein, partial cds.		
ACCESSION	AK024425		
VERSION	AK024425.1	GI:10440358	
KEYWORDS	fts (full insert sequence).		
SOURCE	Homo sapiens adult spleen cDNA to mRNA, clone:as00014.		
ORGANISM	Homo sapiens		

REFERENCE AUTHORS TITLE	REFERENCE AUTHORS TITLE
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 4469)	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 4469)
Ohara,O., Nagase,T., Kikuno,R. and Okumura,K. The nucleotide sequence of a long cDNA clone isolated from human spleen Published Only in Database (2000) In press	Ohara,O., Nagase,T., Kikuno,R. and Okumura,K. The nucleotide sequence of a long cDNA clone isolated from human spleen Published Only in Database (2000) In press
2 (bases 1 to 4469) Ohara,O., Nagase,T., Kikuno,R. and Okumura,K. Direct Submission Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institue Department of Human Gene Research; 1532-3 Yana, Kisarazu, Chiba	2 (bases 1 to 4469) Ohara,O., Nagase,T., Kikuno,R. and Okumura,K. Direct Submission Submitted (24-AUG-2000) Osamu Ohara, Kazusa DNA Research Institue Department of Human Gene Research; 1532-3 Yana, Kisarazu, Chiba

COMMENT

FEATURES
source

gene
CDS

CDS

BASE COUNT
ORIGIN

Query Match	82.9%;	Score 2117.0;	DB 21	Length
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2177; Conservative	1;	Mismatches	0;	Indels 0; Gaps 0;

CTCCAGGCCATGACCAAGTGGTTCTTCCTGGGTGGCCWGAAGCC
CTCCAGGCATGTACTAGATGAGTACCAGAGACCGCCTCTTCTGGT

511 CTTCACTCTCTGCGGGGCTTGAGCCAGGCATGGCCAGATCCCGGAGGCTCTGTGGCCACCG 12

571 CAGCCAGGACAGAGGGAGAGAGTGTGTTCAAGGGAGAGATCCTTTGACACAGATGCGCC 18

QY 631 AACCTGCTGGGCTGCTACAGCCTCACACACGGACCACTGCTAGCCTGTGGCACTGGG 2

```
QY 691 GCTTCCAGCCACCTGTCCTCATCATCTGAGCCACCTGCGGAGCATGTCCTCAC 750
DB 241 GCTTCCAGCCACCTGTCCTCATCATCTGAGCCACCTGCGGAGCATGTCCTCAC 300
QY 751 CTGAGGCTGACATGTCGAAAGTGGCCGCGGCGGTGCTCCAGAGCCACCGCTCC 810
DB 301 CTGAGGCTGACATGTCGAAAGTGGCCGCGGCGGTGCTCCAGAGCCACCGCTCC 360
QY 811 TTTGCGCAGACCTTCATAGAGCGGGAGCTGTACAGGGTCTCATCTGCTGACTTCTG 870
DB 361 TTTGCGCAGACCTTCATAGAGCGGGAGCTGTACAGGGTCTCATCTGCTGACTTCTG 420
QY 871 CGAGAGCCATGATTCCTCGAAGTGAAGTCTCGGCGAGCTGTGCTGCTGCTGAG 930
DB 421 CGAGAGCCATGATTCCTCGAAGTGAAGTCTCGGCGAGCTGTGCTGCTGAG 480
QY 931 CAGAGTCTCTTGACACACCCCGGTTTGTATGAGCGCGCGGATCCCTGAGAACTCTG 990
DB 481 CAGAGTCTCTTGACACACCCCGGTTTGTATGAGCGCGCGGATCCCTGAGAACTCTG 540
QY 991 CAGAGCAATGACAGGTGACTTCTTCTGCGAGAGCTGCTGCGCGGATGCTGAG 1050
DB 541 CAGAGCAATGACAGGTGACTTCTTCTGCGAGAGCTGCTGCGCGGATGCTGAG 600
QY 1051 TCGAACCATGATCTGATGAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 601 TCGAACCATGATCTGATGAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 1111 CGGAGTCTGTCGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
DB 661 CGGAGTCTGTCGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
QY 1171 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 721 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 1231 AAGCCCGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
DB 781 AAGCCCGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
QY 1291 GGGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 841 GGGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 1351 GCGCCAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
DB 901 GCGCCAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
QY 1411 CGCCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 961 CGCCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 1471 AAGGACTACCCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
DB 1021 AAGGACTACCCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
QY 1531 GGGGCGCTGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 1081 GGGGCGCTGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
QY 1591 CAGATCGTGTGAGACCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
DB 1141 CAGATCGTGTGAGACCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
QY 1651 ACTGACTCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
DB 1201 ACTGACTCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1711 GAGGAAGTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1261 GAGGAAGTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
QY 1771 GAGATCTCTGTCAAAAAGGCAATGCTATACGTCGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1321 GAGATCTCTGTCAAAAAGGCAATGCTATACGTCGCTGCTGCTGCTGCTGCTGCTG 1320

DB 1321 GAGATCTCTGTCAAAAAGGCAATGCTATACGTCGCTGCTGCTGCTGCTGCTG 1380
QY 1831 CGGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1890
DB 1381 CGGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
QY 1891 CCAATCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1950
DB 1441 CCAATCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
QY 1951 CGGTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1501 CGGTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 2011 CAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2070
DB 1561 CAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2070
QY 2071 AGCAGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2130
DB 1621 AGCAGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
QY 2131 AGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2190
DB 1681 AGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
QY 2191 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2250
DB 1741 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2250
QY 2251 CTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2310
DB 1801 CTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
QY 2311 CAGCTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2370
DB 1861 CAGCTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
QY 2371 CAGGCTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
DB 1921 CAGGCTGAGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
QY 2431 AACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2490
DB 1981 AACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
QY 2491 TCAGAGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2550
DB 2041 TCAGAGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2550
QY 2551 CTGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2610
DB 2101 CTGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2610
QY 2611 GAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2660
DB 2161 GAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2660

RESULT 6
AX155170 2340 bp DNA linear PAT 22-JUN-2001
LOCUS AX155170
DEFINITION Sequence 1 from Patent W00140278.
ACCESSION AX155170
VERSION AX155170.1 GI:14536655
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2340)
Holloway, J.L. and Foley, K.P.
```

TITLE Human semaphorin 2smf-16
 Patent: WO 0140278-A 1 07-JUN-2001:
 JOURNAL ZymoGenetics, Inc. (US)
 FEATURES Location/Qualifiers

source 1..2340
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2340
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC42673.1"
 /db_xref="GI:1453665"
 /translation="MPSAMALICWLLGILLHGGSSGSPSPVRLRLRYGAMVRK
 /translaltion="MPSAMALICWLLGILLHGGSSGSPSPVRLRLRYGAMVRK
 PSTMMETEFRLLSANRSLFLGPOSLIMLOMYDELDRLFLGDLALSLKD
 QAMPDRETCANFVRLPNNRHLACGTGALOPICALITVGRHEVHLHLEPSV
 ESGRCRCPHSPRPASTEIDELXTGLADFLGRBEMHVRSGSPRALSDOSIL
 HDPRYMARIPENSODNDKYFFFSSTVPSPDGSHNHYSGVRCVNDAGQRY
 LVNKSFTIKARLYVSPGGAETHFDLEDELVLLMKAKSLFVYALFSTYSAVQ
 GFAVCYHMDITVEVFNPGFAHDGPDQWQVGVKVPFPPRGVCAQPDNDTVY
 STADYDEVLQFRAHPLMFVPRHGRPLVATKLAOLHOLVDRVDEADDTVY
 FLGTDGSLVLYTAHQAGSAPSEVEVLELOVFPVPTITEMELSVKRMALVGRN
 LGVADRLHOCETVGTACAECLARDPCAMDGACSTYRSLGRFRERDRIKRN
 ALQICGSOEEVAVLAATVYGTGTHNTELECPKSPQAAVRLVLRPDEBQV
 KTDREVLTERGLFRLSRFDAGTCTTTEHSGSPQVRLAVIVAASDLNLFPP
 EKPEPPARGLASTPPKAMVRIQLQIFANLPVDYCRVWCNGTITCSGCFRS
 RSRKQARCKSMAGLELGRKMSRVHAEHRIFREYEA"

BASE COUNT 406 a 748 c 743 g 443 t
 ORIGIN

Query Match 81.8%; Score 2148.6; DB 6; Length 2340;
 Best Local Similarity 95.1%; Pred. No. 0;
 Matches 2285; Conservative 1; Mismatches 0; Indels 117; Gaps 2;

280 ATGGCCCCCTGGGCTGGGCTGATTTCTGCTGCTAGGGGGGCTCTCTCATGGGGGT 339
 1 ATGGCCCCCTGGGCTGGGCTGATTTCTGCTGCTAGGGGGGCTCTCTCATGGGGGT 394
 340 AGCTCTGGCCCCAGCCCCGGGCGGAGTGTGCCCGGCTGGCTCTCTACCGAG----- 394
 61 AGCTCTGGCCCCAGCCCCGGGCGGAGTGTGCCCGGCTGGCTCTCTACCGAGAGGCC+120
 395 -----ACCTCTGTCT 405
 121 ATGTCCGAAAGCTTCCAGACCATGTGATGAAACATTTTCCAGATACCTCTCTGTCT 180
 406 GCCAACCCCTCTGCATCTTTCTGGGCCCCAGGGCTCCCTGAACCTCCAGGCGATGAC 465
 181 GCCAACCCCTCTGCATCTTTCTGGGCCCCAGGGCTCCCTGAACCTCCAGGCGATGAC 240
 466 CTAGATGATACGAGACCGCTCTTTCTGGGGTGGCTGGAGCGCTCTACTCTCTGGG 525
 241 CTATATGATAGTACGAGACCGCTCTTTCTGGGGTGGCTGGAGCGCTCTACTCTCTGGG 300
 526 CTGAGACGAGGATGGGCGAGATCCCGGAGGTCTGTGTGCCACCCGACGAGAGAGG 585
 301 CTGAGACGAGGATGGGCGAGATCCCGGAG-----329
 586 GAGGAGTGTGTGAAAGGAAAGATCTTTGACAGAGTGGCCCACTTGTGGGGGTG 645
 330 -----GACAGAGTGGCCCACTTGTGGGGGTG 657
 646 CTAGACGCTCACACCGGACCCACCTGTAGCTGTGGGCGACTGGGGCTTCCAGGCCACC 705
 358 CTAGACGCTCACACCGGACCCACCTGTAGCTGTGGGCGACTGGGGCTTCCAGGCCACC 417
 706 TGTGCCCTCATCAGATTGGGCGACGATGTCTCTCACTTGGAGGCTGGAGG 765
 418 TGTGCCCTCATCAGATTGGGCGACGATGTCTCTCACTTGGAGGCTGGAGG 477
 766 GTGGAAGTGGCGGGGGGCGGTGCTTACGAGGCGCCAGCCCTTGTGGCAGACCTTC 825
 478 GTGGAAGTGGCGGGGGGCGGTGCTTACGAGGCGCCAGCCCTTGTGGCAGACCTTC 537

826 ATAGAGGGGAGCTGTACACGAGGCTCTCACTGCTACTTCTCGGGGGGAGAGGACATATC 885
 538 ATAGAGGGGAGCTGTACACGAGGCTCTCACTGCTACTTCTCGGGGGGAGAGGACATATC 597
 886 TTCCGAGTGGAGGCTCTCGGGGAGGCTGCGCTTCCGACTGTGACCGAGATCTTGGCAC 945
 598 TTCCGAGTGGAGGCTCTCGGGGAGGCTGCGCTTCCGACTGTGACCGAGATCTTGGCAC 657
 946 GAGCCCGGCTTGTGATGGGCGCGGAGATCCCTGAGAGTGTGACCGAGGATGACAAAG 1005
 658 GAGCCCGGCTTGTGATGGGCGCGGAGATCCCTGAGAGTGTGACCGAGGATGACAAAG 717
 1006 GTGTACTTCTTCTTCTCGGAGAGGATCCCTGAGAGTGTGACCGAGGATGACAAAG 1065
 718 GTGTACTTCTTCTTCTCGGAGAGGATCCCTGAGAGTGTGACCGAGGATGACAAAG 777
 1066 GTGACCGGCTGGGGCGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 837
 778 GTGACCGGCTGGGGCGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1185
 1126 AATGAGGAGCTTCTCTCAAGGCGAGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 897
 838 AATGAGGAGCTTCTCTCAAGGCGAGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1245
 1186 GAGACCCACTTGTACACGCTAGAGGATGATGATGATGATGATGATGATGATGATGAT 957
 898 GAGACCCACTTGTACACGCTAGAGGATGATGATGATGATGATGATGATGATGATGAT 1305
 1246 CTGAGGAGTACGCGCTGTACAGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 1017
 958 CTGAGGAGTACGCGCTGTACAGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 1365
 1306 GTTACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1077
 1018 GTTACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1425
 1366 CCTGACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1137
 1078 CCTGACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1485
 1426 CCTGACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1197
 1138 CCTGACCAATGAGGAGACATCTGAGAGGATTTTCACAGGCGCTTGTCCACAGGATGAG 1545
 1486 GAGTGTGACGATTTTCCGAGGCGGAGGCTGATGATGATGATGATGATGATGATGATGAT 1257
 1198 GAGTGTGACGATTTTCCGAGGCGGAGGCTGATGATGATGATGATGATGATGATGATGAT 1605
 1546 GCGGCGCTGTCTTGTCAAGACGACCTGTGGCCAGACGACCTTACACAGATGATGATGAT 1317
 1258 GCGGCGCTGTCTTGTCAAGACGACCTGTGGCCAGACGACCTTACACAGATGATGATGAT 1665
 1606 GCGGCGCTGTCTTGTCAAGACGACCTGTGGCCAGACGACCTTACACAGATGATGATGAT 1377
 1318 GCGGCGCTGTCTTGTCAAGACGACCTGTGGCCAGACGACCTTACACAGATGATGATGAT 1725
 1666 GTGTCTCAAGTATGCTCTCTCAGGAGGCGCTGACGAGGCTGACGAGGATGATGATGAT 1437
 1378 GTGTCTCAAGTATGCTCTCTCAGGAGGCGCTGACGAGGCTGACGAGGATGATGATGAT 1785
 1726 GAGGAGCTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
 1438 GAGGAGCTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845
 1786 AGGCAATGCTATAGCTGAGGCTCTCGGGGTGTGGGCGGAGGCTGACGAGGATGATGAT 1557
 1498 AGGCAATGCTATAGCTGAGGCTCTCGGGGTGTGGGCGGAGGCTGACGAGGATGATGAT 1905
 1846 GAGACTTACGAGGACCTGCTGTGAGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1617
 1558 GAGACTTACGAGGACCTGCTGTGAGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1965
 1906 GATGTGCTCTCTGATACCACTACCGGCGGAGGCTTGTGCGGCGGAG 1965

	Indels	Gaps
280 ATGCCCCCTCGGCCTGGGCCAATTTCCTCCTT	15;	2;

RESULT 7	AX155172		
LOCUS	AX155172		
DEFINITION	Sequence 3 from Patent WO0140278.	2337 bp	DNA
VERSION	AX155172		linear
KEYWORDS	AX155172.1 GI:14536657		PAT 22-JUN-2001
SOURCE			
ORGANISM	synthetic construct.		
REFERENCE	synthetic construct.		
AUTHORS	artificial sequence		
TITLE	1 (bases 1 to 2337)		
JOURNAL	Hollway, J.L. and Foley, K.P.		
	Human semaphorin zsmf-16		
	Patent: WO 0140278-A 3 07-JUN-2001;		
FEATURES	Zymogenetics, Inc. (US)		
SOURCE	Location/Qualifiers		
	1. 2337		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/note="Degenerate Polynucleotide sequence of ZSMF-16"		
BASE COUNT	308 a 273 c 468 g 293 t 995 others		

[illegible]

Query Match 24.1% Score 632.4; DB 10; Length 2898;
 Best Local Similarity 57.6%; Pred. No. 2,5e-107;
 Matches 1282; Conservative 1; Mismatches 902; Indels 39; Gaps 7;

```

QY 274 CCCAGCATAGGCCCCCTGGCCCTGGCCATTTGGCTGTCTAGAGGCCCCCTCTCTCCAT 333
DB 364 CGCACATAGGACCGGCGGACACATCTTCACCTTGTCTGAGGTCACCTGTGGAA 423
QY 334 GGGGGTAGCTTGCCCCACCCCGGCGGAGTGTGCCCCCGCTGCGCTCTCTACGGA 393
DB 424 CTCTGACCCCGAGTCACTCCGGAACCCCTCTACCCAGGCTACGCTGTACATATA 483
QY 394 GACCTCTCTCTCCCAACCGCTTGGCATTTCTTGGGCCCCCGGAGGCGCCCTGACTC 453
DB 484 GAACCTTTGGAACTGATAGGACTTCAATATTTCAAAAGCCCCCTTGATTTCTGTATC 543
QY 454 CAGGCATCTACTAGTATGATACGAGACCGCCCTTTTGGGTGGCTGGAGCCCTC 513
DB 544 CATACATCTCTGTGATGATGATCAAGACGGCTTTTGGGAGGCGAGAGACCTGTCT 603
QY 514 TACTCTCTCGGCTGGACGAGCATGGCCAGATCCCGGAGTCTCTGTGGCCACCGCAG 573
DB 604 TATTCCTGTAATTTGAAACGAGTCACTGACGCTACAGAGATATCTGGCCGAGACA 633
QY 574 CCAGACAGAGAGGAGAGTGTGTTCGAAAGGAGAGATCCCTTGCAGAGTGTGCCAAC 633
DB 664 GCAGTAAGGTAAGAAATGATGATGAAAGAAAGACCC---AAATGATGTGCCAAT 720
QY 634 TTCTGCGGGGTCTACAGCCCTCAACACCGACCCACTGCTACCTGTGGACCTGGGCG 693
DB 721 TATATCCGGGTTTGCATCTACACAGGACACCTTCTGTGACCTGTCTGTAGAGT 780
QY 694 TTCCAGGCAACCTGTGCTCTCATACAGTTGGCCACCGT---GGGAGCATGTGCTCAC 750
DB 781 TTGTATCTACACTGTGCTCTCATACAGTGTGGGACCATTTACAGAGAACCCCTGTTAC 840
QY 751 CTGAGCTGTGCACTGTGAAAGTGTGGCGGGGCGGTGCTCTCAAGAGCCAGCCCTCC 810
DB 841 CTGGAATCTACACAGATCTGAGAGAGGAGGACATGTCTTTTGGACCCCACTCTCC 900
QY 811 TTTCAGACACCTTCTTAACAGGGGAGCTGTACAGGGTCTCACTGTCTCTCTGCTGG 870
DB 901 TTTGTCTCAAGCTTATGAGATGAGCTGTGTGTGCTGACTCTACAGTATATTTGGGCG 960
QY 871 CGAGAGGCATATCTTCCGAAAGTGAAGTCTCGGCGACGCTGCGCTCGCATCTGAC 930
DB 961 AAGAGCTGCGCATCTTCCGAGCATGGGGAATTAGGCCATATTCGACTGAGCATGAC 1020
QY 931 ---CAGAGTCTCTGACGACCCCGGTTGTGATGCGCCCGCGATCCCTGAGAACTCT 987
DB 1021 GATGAGCGGCTCTGAAAGAACAAATTTGTAGTTTCATATATGATTCCTGATACGA 1080
QY 988 GACGAGGACATGACAGGTGTACTTCTTCTCGGAGAGGCTGCCCTGCGCCGATGTGT 1047
DB 1081 GACCGAATGACAAATAAATTTCTTTTCTTACTGAGAGCGGTGGAAGCGGGGAAC 1140
QY 1048 GGTCTGACCATCTCACTGTGACCGCGGTGGGCGGCTGTGCTGTAAGTANTCTGGGGC 1107
DB 1141 AAGCCCAACAGATTA---CACCGAGTGGGCGGCTGTGCTAAGTACATGGAGGA 1197
QY 1108 CAGCGGCTGTGTGAACAAATGAGACATTTCTCAAGCGCAGCTGTGCTGCTGGT 1167
DB 1198 CAGAGATCTGTGTGAACAGTGAAGCACTTCTTAAAGCGCGCTGTGTCTCAAGT 1237
QY 1168 CCGGCGCTGTGTGTCGAGACCCACTTTCAGCAGTGAAGATGTGTCTCTCTGTGG 1247
DB 1258 CCGGGAATGAATGAGATGACACATATTTGACGACTAGAGATGTGTCTTACTCCG 1317
QY 1228 CCCAAGGCCGGAAGAGCTGTGAGTGAAGCGCTGTTCAGACACCTGACAGCGCTGT 1287
DB 1318 ACCAAGATCTCAAGAACTCACTGATATTTGAGCTGTATATCTACACAGATATTTT 1377
QY 1288 CAGGCTTTCGCGCTGTGTGTACCATGTGACAGACATCTGGAGGTTTCAAGGCGCC 1347

```

```

DB 1378 AGAGGCAATCTGTATGTATGTATCATCATGTCAAGTATCCGGGAAGCCTTTATGGCCA 1437
QY 1348 TTTCCCAACGAGATGAGGCTCTGACACAGTGGGGGCGCTATGGGGCAAGGTGCCCTTC 1407
DB 1438 TATGCTCATATAAAGAGCGCTGTAATACACTGTCTCAATATGAAGAAAGTCCCTTAC 1497
QY 1408 CTTGCGCTGTGCTGTGCCCCAGCAAGATGACCGACAGACCGAGGAGGCTTTTGGGAGC 1467
DB 1498 CCAAGGCTGTGCTGTGCGACGCAAAAGTAAAC-----GGAGGCAATATGGAAAC 1548
QY 1468 ACCAAGACTTACCAGATGAGTGTGCTGATGCTTGGCCGAGCCACCCCTCATGTCTGG 1527
DB 1549 ACCAAGATTTACCCGATGATGACCCATCCGCTTCCAGAGATGCAATCTTAATATACG 1608
QY 1528 CCTGTGCGGCTCTGACATGCGCGGCTCTCTTGTCAAGACCCACTGTGGCCAGCAGCTA 1587
DB 1609 CCCATATAACCTGTTCAATAAATAACCAATACTGTGTAATAAAGATGAAATACAACTG 1668
QY 1588 CACGATCTGTGTGAGCCGCTGTGAGGACAGAGATGGAGCTACAGATCTATTTCTG 1647
DB 1669 AGGCACTTGGCTGTGATCGGTGGAAGCGAGATGGCAGTATACGCTTATTTATTT 1728
QY 1648 GGAAGTACTGAGGCTGTGCTCAAGTCAATGCTCTCCAGGCAAGGCGCTCAGCTGAA 1707
DB 1729 GGGACAGACAGAAATGTGCTGAAGTATATCAATTTACAAACCAAGAAACAGGTGG 1788
QY 1708 CTTGAGAGATGTGTGAGAGAGCTCCAGGTGTTTAAGTGCACAACTCTACCCGAA 1767
DB 1789 AVGGAAGATCTATCTGAGAGAACTTCAAAATATTCAGAGATCCAGCCCTATCTTCT 1848
QY 1768 ATGAGATCTCTGTCAAAAGGCAATGCTATACGTGGGCTCTGCGCTGTGGTGGCCAG 1827
DB 1849 ATGGAATTTCTTCAAAAGAGACAAACGCTTTACATTTGATCAAGCTCTGTGCTGACAA 1908
QY 1828 CTGCGCTCACCAATGTAGACTTAACGCACTGCTGTGAGAGTGTGCTGCTGGCCGG 1887
DB 1909 GTTCAGATCTCATACAGTGCACATGTATGAGCAAGTGTGCTGTGCTGCTGCTGCTG 1968
QY 1888 GACCCATCTGTGCTGAGATGTGCTCTGTACCCACTGTACCCGCTGCTGCTGCTGCTG 1941
DB 1969 GACCGTACTGTGCTGAGATGAGCAATATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1968
QY 1942 GGCAGGCGGCTTCCGCGGAGAGACATCCGCGAGCGCAACCTGCTGAGTGTGCTG 2001
DB 2029 GCAAGAGAGAGTTCGCGAGGAGGCTTCCGAGTGTGCAAGCCCGCCCAACAGTCTTT 2088
QY 2002 GGCAGAGCTCAGAAAGAGAGGCACTGTGTGCAAGCCCACTGTGCTTATGGCAGC 2061
DB 2089 GGACGCAATTTGTGAGAGAGCGGCTGTGAGAGCACTGAAGAGAGGCTGTATGGCATA 2148
QY 2062 GAGCAATATGACACTTCTGAGATGCTGCTCCCAAGTGTCCCAAGCTGTGCTGCTG 2121
DB 2149 GAGAGCAACAGTACTGTGTGAATGACACCCGCGATCTCAAGCAAAATCTATCTGG 2208
QY 2122 CTCTGAGAGGAGGAGGAGTGAAGGCGCTGACAGGTGAAGACGAGAGAGAGTCTTG 2181
DB 2209 TTGTGACAGAGAGGAGCGCAGTAAAGAAAGAGGTGAAGAGCGATGACAGAGTGTCT 2268
QY 2182 CACAGGAGCGGCGGCTGTGCTGCGAGGCTTACCCCTTTCAGTGTGCGGAGCTTACAC 2241
DB 2269 AAGATGAGCTGTGCTGTCTCTCTCAAGATGAGTCAAGTCAAGGAGGAGCTATTTT 2328
QY 2242 TGCACACTGTGAGATGAGCTTCTCCAGAGTGTGTGCTGCGCTGTGCTGTGATTT 2301
DB 2329 TCCGACAGAGTGAACCAATTTTGTCTCAATGTGTGCTTAAATCACTGTGAGGTGTCT 2388
QY 2302 GTGGCTCAAGCTGTGAGCAACCTTCTCCGAGC-----CAAGCCAGAG 2349
DB 2389 GAAAGCATTAAGTGAAGGAGCTTTCATAGAGACATGAAGAGAAAGACATTCACAAG 2448
QY 2350 CAGCCCGGAGCGGAGGAGCTGTGCTTCCACCCCAAGGCTGTGTACAGAGATCT 2409

```

Db 2449 ATGCCTGCGCCCTTCCCTTAGCGGTATGTCCTCAGGGGACAAAACCGTGTACAGAAATTC 2508

QY 2410 CTCAGCTCATTTGGCTTTCGCCAACCCCGCGGTGATGACTGTGAGCGCGGTGGT 2509

Db 2509 TTGCAGCTGATTTGGCTTTCGCCAACCCCGGTGATGACTGTGAGCGCGGTGGT 2568

QY 2470 TGCA 2473

Db 2569 TGTG 2572

RESULT 9

LOCUS MUSEMH 2328 bp mRNA linear ROD 20-MAY-1998

DEFINITION M. musculus mRNA for semaphorin H.

ACCESSION 280941

VERSION 280941.1 GI:3150195

KEYWORDS semaphorin H.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 2328) Christensen C.R., Klingelhofer, J., Tarabyskina, S., Hulgaard, E.F., Kramarov, D. and Lukanidin, E. Transcription of a novel mouse semaphorin gene, M-semah, correlates with the metastatic ability of mouse tumor cell lines Cancer Res. 58 (6), 1238-1244 (1998)

REFERENCE 2 (bases 1 to 2328) Christensen, C.C. Direct Submission Submitted (11-Oct-1996) Christensen C. C., Danish Cancer Society, Molecular Cancer Biology, Strandboulevarden 49, Copenhagen, Denmark, 2100

REMARK 3 (bases 1 to 2328) revised by [3]

REFERENCE 3 (bases 1 to 2328) Christensen, C.C. Direct Submission Submitted (14-APR-1998) Christensen C. C., Danish Cancer Society, Molecular Cancer Biology, Strandboulevarden 49, Copenhagen, Denmark, 2100

COMMENT On May 22, 1998 this sequence version replaced gi:1619599.

FEATURES

Source

1. .2328

/organism="Mus musculus"

/strain="Balb/c FC3H"

/db_xref="taxon:10090"

/cell_line="66c14"

1. .2328

/codon_start=1

/product="semaphorin H"

/protein_id="CAB02590.1"

/db_xref="GI:3150196"

/db_xref="SPTREMBL:P70275"

/translation="MAPAGHLLTLLMGLHLLMTPOHSANPSYPRRLSHKELLFLN
RSTQSPGIGLIDHTMLDDEYERLIVGGGRDLYSLNLEKVDGREGYTWSTAYKV
ECTKMGKADNACANTRYLVHNYRTLLCAAGAFRNGHSHSEPLEHED
HSEKRGRCPPDRSSFTLVNENELFAGISYDWRDCAFLRSMKSLHITENDG
BRLEKRFVGSYMLPDEBDNDKMLFEPEKYLEENNAHTLYTVGGLAYNDG
ORLVNWKSTELKAVLCVSGVANGIDTYPDELDVTLPPRDPKSGCAKVGKRG
IFRGHACVYHMSIREAFNGPIVAKRBPETVHSLYSGKVPKPGSCAKVNGKRG
TKDVPDAIRFARIDPLMOPIKRVHKKPLVTDGKYNLRQALRVRAEDGOYV
LFGDGTGLVLTIVYNOETEMMEVLELELQIFKDPAPLISHEISKOOLYISGA
SAVQVRFHCHDMYGSACADCLADPYCAMDGISCSRYPLTGAAHARRRODVRG
NAAOCFGQDFGADLDRTEERLAGEISNTLSECTPRSLQAVLWFOGRVYKE
EYKIDRVVAKMLDGLITRVKRSADGATTECOVEHNFVTRKLTLEVEPEHVEGMF
HKDHEERHAKFCPLSLMSGCTPKWKEFLQILGYENFORVEYCEKWCIDRKRR
KLMSPSKWRYANPODKRLRSKAERFRLRHLLS"

BASE COUNT 665 a 542 c 595 g 526 t

Query Match 24.0%; Score 629.6; DB 10; Length 2328;

Best Local Similarity 57.6%; Pred. No. 8,4e-107; Matches 1278; Conservative 1; Mismatches 900; Indels 39; Gaps 7;

QY 280 AAGGCCCCCTCCGCGCTGGGCCATTTGCTGCTGCTAGGGGGCCCTTCATGAGGGGT 339

Db 1 AAGGACCGGCGGACACATCCCTCACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60

QY 340 AGCTTGGCCCGACGCCCGCCCAAGTGTGCGCCGCTGCGGCTCTCTACCGAGCTC 399

Db 61 ACCCGAGTCTACTCCGCGAACCCTCTCTACCCAGGTACCGCTGTACATTAAGAACTT 120

QY 400 CTGCTCCCAACCGCTCTGCGCATCTTGTGGGCCCCCAGGCTCCCTCACTACCGAGCC 459

Db 121 TTGAACTGATGAGACTTCATATTTCAAGCCCTTGTGGGCTTTCTTATGATCATACA 180

QY 460 ATGTACTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240

Db 181 ATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240

QY 520 CTGCGGCTGACACGAGGATGCGGATGCGGAGGAGTCTGTGGGCTCCGACCGACGAGA 579

Db 241 CTGAACCTTGAACGAGTACGACGCTACAGAGATATATATATATATATATATATATAT 300

QY 580 CAGAGGAGAGAGTGTGTGCAAGAGAGAGATCTTGAAGAGTGTGCGCAATCTGCTG 639

Db 301 AAGGTAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 357

QY 640 CGGAGTCTACACCTTCACACACCGGACCTCTGTAGCTGTGCTGTGCTGTGCTGTGCT 699

Db 358 CGGCTTTCATCTACATACACAGACGACACCTCTGTGCTGTGCTGTGCTGTGCTGTG 417

QY 700 CCCACTGTGCTCTATGACACGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 756

Db 418 CCACACTGTGCTCTATGACACGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 477

QY 757 CTTGCAAGTGTGGAAGTGTGCGGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 816

Db 478 TCACACAGATCTGAG 537

QY 817 AGCACTTCTATACAGGAG 876

Db 538 TCCACGCTATGAG 597

QY 877 GCCATGCTTCTGCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 933

Db 598 TCGGCACTTCTGCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 657

QY 934 AGTCTTGTGCAAG 993

Db 658 CGGCTCTGGAAG 717

QY 994 GACATGACAGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1053

Db 718 GATGACACAAAT 777

QY 1054 AACCATGCTACTGTAG 1113

Db 778 CACACAGATCTA---CACCGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 834

QY 1114 GTGCTGTGCAAG 1173

Db 835 ATCTGTGTAACAG 894

QY 1174 CTTGTGTGTGCGAG 1233

Db 895 ATGAATGAGATGACACATCTTGTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954

QY 1234 GCGGAG 1293

Db 955 GATCTGAAG 1014

QY 1294 TTGCGCTGTGTGTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1353

Db	1015	ATGCTGTATGTGTGATTCACANGTCAAGTATCCGGGAAGCCCTTAAATGAGCCATATGCT	1074
QY	1354	CACCGAGATGGCCCTCAGACACAGTGGGGGCCCTATGAGGGGCAAGTGGCCCTCCCTGGC	1413
Db	1075	CATTAAGAAAGGCCCTGAATATCTTCCCTACTGGTCACTATATGAAGGAAAAGTCCCTACCCAAAG	1134
QY	1414	CGTGGCGTGTGGCCCAAGAGTACCAGACAGCCAGGACGGCCCTTTTGGCAGCAAG	1473
Db	1135	CGTGTGTTCCGTGTCCAGCAAAATGAAC-----GGAGGCAAGATGAAACCAACAA	1185
QY	1474	GACTACCCAGATGATGAGTGTGTCGATGTTTGGCCGAGCCACCCCTCATTTGTGGCGTGTG	1533
Db	1186	GATTACCCCGAATGACGCCATCCGGTTCCGGAAGATCGATTCCTATATATACAGCCATA	1245
QY	1534	CGGCTCTGCAGATGGCCCGCCCTGCTCCTTCTCAGAACCCCACTGGCCACAGCTACACAG	1593
Db	1246	AAACCTGTTCAATAAAAAACCAATACTGGTAAAAAAGATGGAATAATCAACTGAGGCA	1305
QY	1594	ATCTGTGTGACCCGCTGTGAGGACAGAGATGGAGCTACGATGTTCATTTTCCTGGGACT	1653
Db	1306	CTTGGCGGTGATCGGGTGGAGCGAGGATGGCCAGTATGACGTCTTATTTATTTGGACA	1365
QY	1654	GACTCAGGGTGTGTGCTCAAGTCAATGCTTCCAGGCAAGGGGGCTCAGCTGAACCTGAG	1713
Db	1366	GACACAGGAATTTGTCTTAAGTAATCACAATTTTACAACCAAGCAACAGATGAGATGGAG	1425
QY	1714	GAACTGTGTGAGAGGAGCTCAGGTTTAAAGTGGCAACACTATACCCAGAAATGAG	1773
Db	1426	GAAAGCATTTCTAGAGGAACCTCAAAATATTCAGAGATCCACCCTTATCATTTTATGGA	1485
QY	1774	ATCTGTCAAAAGGCAAAATGCTTACGTGGCTCTCGGCTGGTGTGGCCAGCTGGCG	1833
Db	1486	ATTCTTCAAAAGACACAACGCTTTACATTGGATTCAGCCCTCTGCTGTGACACAATGCAG	1645
QY	1834	CTGCACCAATGTGAGACTTACGGCAGCAGCTGTGACAGAGTGGCTGGCCCGGAGACCA	1893
Db	1546	TTCATCACTGCGCAATGATATGACAGAGCTTGTGCTGACTGCTGGCTCGAGACCCG	1605
QY	1894	TACTGTCCCTGGGATGGTGGCTCCTTACCCACTACCGCCCA-----GCCTTGGCAAG	1947
Db	1606	TACTGTCCCTGGGATGGCATATCTCTCTCCAGTACTACCAACAGGTGCACACCAAAAG	1665
QY	1948	CGCGCTTCCGCCGACAGACATCCGACAGGCAACCTCGCCCTGACAGTCCGTGGGCCAG	2007
Db	1666	AGAGAGTTCGCGAGGACAGACGTTGGCATGGCMAACGCCGCCCAACATGCTTGGCAAG	1725
QY	2008	AGCAGAGAAGAAAGGCATGTGGACTTGTGGCAGCCACCATATGTCTTACGGCAGGAGCAG	2067
Db	1726	CATTTGTTGGAGACCCGTTTGGACAGAGACTGAAGAAGAGGCTGGCTTATGGCATGAGAGC	1785
QY	2068	AATAGCACCTCTCTGGAGTGCCTGCGCAAGTCTCCCAAGCTGGCTGTGGCGTGGCTCTTG	2127
Db	1786	AACAGTACTCTTGTGGAATGACCCCGGATCACTACAGCAAAAGTCATGTGGTTTGTGA	1845
QY	2128	CAGAGGCCAGGGGATGAGGGGCTTACCAGTGAAGACGACGAGACGATCTTGCACACG	2187
Db	1846	CAGAAAGGACGCGATGAAGAAAAGAGGTGAAGAGGATGACAGAGTGTCAAGATG	1905
QY	2188	GAGCGGGGCTGTCTTCCGACAGGCTTGGCGGTTTCGATGGGGGACCTTACACTGCACC	2247
Db	1906	GACTTGGGCTGTCTTCTCTCAGAGTACGCAAGTCATATGACAGGACCTTATTTTTCGAG	1965
QY	2248	ACTGTGGAGCATGGCTTCTCCCAACATGTGGTCCGCTGGCTCTGTGTGATTTGGGCC	2307
Db	1966	ACAGTGAACACAAATTTTGTTCATAATCTGTGTAAATTCACTTGGAGGTGTGCAAGAG	2025
QY	2308	TCACAGCTGACAACCTGTTCCTCCGAGAG-----CAAAGCAGAGGAGCC	2355
Db	2026	CATAAAGTGGAGGCGCATTTTCATTAAGACCATGAAGAGAAACATCAAGATGCCC	2085
QY	2356	CCAGCCCGGGGAGGCCCTTGCTTCACCCCAACCCAGAGCCTGTGATCAAGACAATCTGCAG	2415
Db	2086	TGCGCTCCCTTAAAGCGGATGTCTCAAGGGGACAAACCGGTTACAAAGGAATTTTGGCAG	2145

QY	2416	CTCATTTGGCTTGGCCCAACCTGCCCGGCTGGTGAATGACTCTGTGACCGCGTGTGTGCA	2473
Db	2146	CTCATTTGGCTTGGCTTGGCCCAACCTGCCCGGCTGGTGAATGACTCTGTGACCGCGTGTGTGCA	2203
RESULT 10			
HSU28369			
LOCUS	HSU28369	2919 bp	mRNA
DEFINITION	Homo sapiens semaphorin V mRNA, complete cds.		
ACCESSION	U28369		
VERSION	U28369.1		
KEYWORDS	GI:974283		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 2919) Wei,M.H., Latif,F., Bader,S., Kashuba,V., Chen,J.Y., Duh,F.M., Zbar,B., Minna,J.D., Cell,L., Kuzmin,I., Zbarovskiy,E., Klein,G., Sekido,Y., Lee,C.C., and Lerman,M.I.		
TITLE	Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: progress toward the isolation of a lung cancer TSG		
JOURNAL	Cancer Res. 56 (7), 1487-1492 (1996)		
MEDLINE	96181293		
REFERENCE	2 (bases 1 to 2919)		
AUTHORS	Albanesi,J.P., Lee,C.C., Lerman,M.I., and Minna,J.D.		
TITLE	Human semaphorin A(V) and IV reside in the 3p21.3 small cell lung cancer deletion region and demonstrate distinct expression patterns		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (9), 4120-4125 (1996)		
MEDLINE	96210603		
REFERENCE	3 (bases 1 to 2919)		
AUTHORS	Sekido,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1995) Yoshihata Sekido, Simmons Cancer Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8590, USA		
FEATURES			
source	Location/Qualifiers		
	1..2919		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="3"		
	/map="3p21.3"		
	236..2485		
	/codon_start=1		
	/product="Semaphorin V"		
	/protein_id="AA09138.1"		
	/db_xref="gi:974284"		
	/translation="MGRGAAAVIPIGLALLMAYGLGSAAPSPPRILRISFOELQANHGL OTFLEIETCCYOALLVDEERGRLPFAENHNVASINTDNISKRAKLAMPADVEMWEC NMARCKDGTGECMNPVRLILVAVNTHLLAGTGAHFHTCAVFGVRAAEPLVLRDPGR IEDKSGSPDPRHNRASVLYVGEELISGAADMGDETFIRSLQGRVSOILRNDGSG RLINPRKVKVPRVPESENPDDEKITYFEFFETAVEAPALGRISVRVQILRTPHDSR RSIYNKRTTTLKRLVCSVPGVEGDHDFQLODVELLSRDHRTPLIYAATGSSSTIF DPEPDVIVAFRNPILMYNSVILPTGGRPLFLOYGANTFQIADRVAAADHYVTLT GTGDVILAKTVISPKGSRSAEGLLEELHVEDSAATVSMQISKRQILYASRSV AQLALHRCANHAGVETCECCALRPPYCAMGVACTCPQSAKRPRRQDVNRNGDSTLC SGGSRPALLENHKEVFEVGSSAPALFCEPSDIAQREMTFQRAGVTAHVOLAEBRT TAGGLLRRLRRDGSVYLCAVDEGFTPLRLSLHVLSTQVAREEAAPAP PGKILWYRDLQIVFEGGGGSANSLMCRPPALOSIPLLESRKRGRRNRTTAPPRAE RGRSRSTHW"		
CDS			
BASE COUNT	531 a	926 c	967 g 495 t
ORIGIN			
Query Match	23.1%	Score 606.8:	DB g: Length 2919;
Matches 1179; Conservative	59.2%;	Pred. No. 1.3e-103;	
	1; Mismatches 773;	Indels 39;	Gaps 7;

Thu Oct 10 09:34:59 2002

us-09-813-290-1.rge

Page 17

[illegible][illegible]

Db 600 GGGAGGAGGATATCTAGGGGTTGGCAGACCTCATGGAGAGATCTTACCATCTTT 659
 QY 889 CGAAGTGGAGGATCTCGCCAGAGCTCTGCGTTCGACTCGACAGAGTCTTTCAC 659
 Db 660 GCGAGCTAGGGAGAGAGCTCCAAAGTCTCCGAAAGAGAGAGAGAGAGAGAGAG 945
 QY 946 GACCCCGGTTTGTATGAGGCGGCGGATCCCTGAGAACTCTGACAGAGAGAGAG 719
 Db 720 GAGCCCAAGTGTCAAGGATATTTTGGATCCCGAGAGAGAGAGAGAGAGAGAG 1005
 QY 1006 GTGTACTTCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
 Db 780 ATCTACTTCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065
 QY 1066 GTCAAGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
 Db 840 GTGTCCGCGGTTGGCAGATCTCCGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1125
 QY 1126 AAGTGGAG 899
 Db 900 AAGTGGAG 1185
 QY 1186 GAGACCCAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 956
 Db 957 GAGACCCAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1245
 QY 1246 CTGAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1016
 Db 1017 CCGTGTCTTATGCGTCTTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
 QY 1306 GTGTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1076
 Db 1077 GTGTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365
 QY 1366 CTGTAG 1136
 Db 1137 CCATGAG 1425
 QY 1426 CCGAG 1195
 Db 1196 CCGAG 1485
 QY 1486 GAGGAG 1243
 Db 1244 GATGTATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545
 QY 1546 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1303
 Db 1304 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605
 QY 1606 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1363
 Db 1364 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1665
 QY 1666 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1423
 Db 1424 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1725
 QY 1726 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1483
 Db 1484 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785
 QY 1786 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1543
 Db 1544 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
 QY 1846 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1603
 Db 1604 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1905
 QY 1906 GCGGCGGCTGTCTTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663

Db 1664 GACGGGTTGCGGTGACAGCGCTTCCAGCCAG---TGCCAAAGAGGGGTTCCGGGGCA 1720
 QY 1966 GACATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720
 Db 1721 GACATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2025
 QY 2026 GTGGAG 1780
 Db 1781 CTGAG 2085
 QY 2086 TGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1828
 Db 1829 TGCGAG 2145
 QY 2146 GCGGAG 1888
 Db 1889 GCGGAG 2205
 QY 2206 GCGGAG 1948
 Db 1949 GCGGAG 2265
 QY 2266 TCCAG 2008
 Db 2009 ACAG 2325
 QY 2326 TCCGAG 2068
 Db 2069 GCGGAG

RESULT 13
 MM293948
 LOCUS M.musculus mRNA for semaphorin Hv (3988 bp). Linear ROD 17-JUL-1998
 DEFINITION M.musculus mRNA for semaphorin Hv (3988 bp). Linear ROD 17-JUL-1998
 ACCESSION 293948
 VERSION 293948.1 GI:1934966
 KEYWORDS semaphorin H.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1. Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Christensen, C.
 TITLE Submitted (10-APR-1997) Christensen C., Strandboulevarden 45, Copenhagen, Denmark, 2100
 REFERENCE 2. (bases 1 to 3988)
 AUTHORS Kramers, D. and Lukanidin, E.
 TITLE Transcription of a novel mouse semaphorin gene, M-semah, correlates with the metastatic ability of mouse tumor cell lines
 JOURNAL Cancer Res. 58 (6), 1238-1244 (1998)

FEATURES

CDS
 gene
 610..2943
 /gene="msemah"
 /note="msemah"
 /note="predicted protein. The coding sequence of the msemah transcript is identical to the submission with Acc. number 280941."
 /codon_start=1
 /product="semaphorin H"
 /protein_id="CAB07988.1"
 /db_xref="GI:1934967"
 /db_xref="SPRMBL:P70275"

Db 2389 ATAGAGACAAACAGTACTCTGTGGAATGCACCCCGCTATCATTACAGCAAAAGTCATC 2448
 QY 2119 TGGCTCTGCAGAGCCGAGGATGAGGGCCCTGACACAGTGAAGACGAGCAGCAGTC 2178
 Db 2449 TGGTTTCTACAGAAAGGAGCGGACGCTAAGAAAGAGAGGTGAAGACGATGACAGAGTT 2508
 QY 2179 TTGCACACGAGCGGGGCTGCTGTCCGACGGCTTACCCGCTTTCAGATGCGGACCTAC 2238
 Db 2509 GTCAAGATGAGCTGGGCTGCTCTCTCTCAGAGTACGACAGTACAGTACGAGGACCTAT 2568
 QY 2239 ACCGTGACACCTCTGAGACATGCTCTCCAGAGCTGTGCTGCGCTGCTGCTGCTGCTG 2298
 Db 2569 TTTTGCACAGAGTGAACCAATTTTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 2298
 QY 2299 ATTGTGCTCTCAGAGCTGAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2628
 Db 2629 GTCAAGAGCATTAAGTGAAGGAGGATCTTTCATTAAGACCATGAAAGAGAAACATCAC 2688
 QY 2347 GAGAGACCCCGACCCGCGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2406
 Db 2689 AGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748
 QY 2407 ATCTGACACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2463
 Db 2749 TTTTGCACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808
 QY 2464 GTGTGCTGCA 2473
 Db 2809 GTGTGCTGTA 2818

RESULT 14
 MM293947 4466 bp mRNA linear ROD 17-JUL-1998
 LOCUS M.musculus mRNA for semaphorin H (4466 bp).
 DEFINITION 293947
 ACCESSION 293947 GI:1934964
 VERSION semaphorin H.
 KEYWORDS house mouse.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4466)
 AUTHORS Christensen, C.
 TITLE Direct Submission
 JOURNAL Submitted (10-APR-1997) Christensen C., Danish Cancer Society, Molecular Cancer Biology, Strandboulevarden 49, Copenhagen, 2 (bases 1 to 4466)
 REFERENCE Christensen, C.R., Klingelhoefer, J., Tarabykina, S., Hulgaard, E.F., Kramerov, D. and Iukanidin, E.
 TITLE Transcription of a novel mouse semaphorin gene, M-semah, correlates with the metastatic ability of mouse tumor cell lines
 JOURNAL Cancer Res. 58 (6), 1238-1244 (1998)
 FEATURES
 source
 Location/Qualifiers
 1..4466
 /organism="Mus musculus"
 /strain="Balb/c/cf3H"
 /db_xref="taxon:10090"
 /cell_line="66c14"
 610..2943
 /gene="msema"
 610..2943
 /gene="msema"
 /note="predicted protein; The coding sequence of the msema transcript is identical to the submission with Acc. number 280941"
 /codon_start=1
 /product="semaphorin H"
 /protein_id="CA07987.1"
 /db_xref="GI:1934965"
 /db_xref="SPTREMBL:P70275"

/translation="MAPAGHITLTLKSHLELTPGHSANPSYARLPLSHKELELNL
 GLQYKAPKAPLIDTLTLDELTOEELFVGSDLYVSLNEFVSQDREIYPSYAVY
 EECITKMGDANECANVITVLHNYNTHLTLTATGAFDHCAPFIVHGSEEDPLHS
 HSEFGRGCEPDPNSSFVSLVGNELFAGLSYWDGSDAIFSRMKGITRTREHD
 ERLLEKPFVSSYMLPDNEDDNKMYFFEFELKLENNATHLHSGRLCIVNDWG
 ORILVKNSTFLKALVCSVPGMNGIDYFDELPLDPLTRDKNFVIGLFNTTS
 IFRHANCVYHMSIREAFNGPYAHKEEPYHKEEYLSGRPYPRPSCASVNGKYV
 TNORLPDAIFRPHPLMOPRIPVYKPLVATDGRVNLRLDADVVEADGQYD
 LFIDGDSIVLKVITTYNOTEMEEVLELEIDFDPAPILISMEISRQDLYS
 ASAYAVRFHHCMDYGSACDCCARPYCAMDEISCSRYPTGAEHRRFRDVRH
 GMAOCCGCGFVDAIDRTREELAYIESNSILLETPLSLAKTIVLFLQKRDYR
 EGVTTDDRVYAMDGLFLFVRYKSDACTYFCQYEHNFVTRVITLFEVHEKVGCM
 FKHDEERHMKMCPPLSGMSQCTKMEYKEPFLQIGYSSKPRVBEYCKWCTDK
 RKLKMSPSRKKMYANPOEKRLRSKAEHFRPLRHLLS"
 BASE COUNT 1289 a 950 c 1024 g 1203 t
 ORIGIN

Query Match 21.7%; Score 570.8; DB 10; Length 4466;
 Best Local Similarity 56.88; Pred. No. 5,9e-96;
 Matches 1266; Conservative 1; Mismatches 918; Indels 45; Gaps 10;

QY 274 CCAGCATGCGCCCTCGGCGCTGCGCCATTGCTGCTGCTGAGGGGCTCTGCTCAT 333
 Db 604 CGCACATGCGACCGCGCGGACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663
 QY 334 GGGAGTACTGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 393
 Db 664 CTGCGACCCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
 QY 394 GACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
 Db 724 GAACCTTTCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 783
 QY 454 GAGGCACTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 513
 Db 784 CATACATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
 QY 514 TACTCTGCGCTGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 573
 Db 844 TATTCCTTAACCTTGAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 903
 QY 574 CCAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 633
 Db 904 GCAGTAAGAGTAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 634 TTGCGGCGGCTGCTGACGCTGACACCGGACCGGACCGGACCGGACCGGACCGGACCG 693
 Db 961 TATATCCGGGTTTTCATCACTACACAGACACACCTTTCGACCTGCTGCTGCTGCTGCTG 1020
 QY 694 TTCCAGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 750
 Db 1021 TTGTATCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 QY 751 CTGAGCCTGCGCATGTGGAAGTGCGCGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 810
 Db 1081 CTGAGGATACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 QY 811 TTGCGACACCTTTCATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 870
 Db 1141 TTTTGTGCGACGCTAGTTGGGAATGAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 QY 871 CGAGAGCCCATGATCTTCCGAAGTGAAGTCTGCGCCGCTGCTGCTGCTGCTGCTGCTG 930
 Db 1201 AGGAGCTGCGGATCTTTCGACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 QY 931 --CAGAGTCTTTCGACACACCGGCTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 987
 Db 1261 GATGAGCGGCTCTGGAAGACCAAAATTTTGAAGTATATATATATATATATATATATAT 1320
 QY 988 GACCCAGAAATGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1047
 Db 1321 GACCCGAGTACACAAATATCTTTTCTTACTGAGAGGCGCTGAGAGCGGAGAAC 1380

1048 GGCTCGAACCATCTACTCTGACGCCCGGCGCTGCTGCTGATGATGCTGGGGC 1107
 1381 AACGCCACACGAT--CTTACACCCGAGTGGAGCGGCG-TGTGCTGATGATGATGAGGA 1437
 1108 CAGCGGTGTGTGTGAACAATGAGCAGCTTCTCTCAAGCCAGGCTGTGTCTGTCTG 1167
 1438 CAGAGATCTGTGTGAACAAGTGGAGCAGCTTCTCTCAAGCCAGGCTGTGTCTGTCTG 1497
 1168 CCGGCGCTGTGTGTGCGAGAGCCCTTACAGCTTACAGAGATGTGTCTGTCTGTCTG 1227
 1498 CCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1557
 1228 CCGGAGCGCGGAG 1287
 1558 ACCAGATCTGTGTGAACAAGTGGAGCAGCTTCTCTCAAGCCAGGCTGTGTCTGTCTG 1617
 1288 CAGGCGCTGTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1347
 1618 AGAGCGCATGT 1677
 1348 TTTGCGCCAGCAGATGGGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407
 1678 TATGCTCTAATAAG 1737
 1408 CCGGCGCTGTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1467
 1738 CCGAGCGCTGT 1788
 1468 ACCAGAGCTGT 1527
 1789 AACCAAGATGT 1848
 1528 CCGGCGCTGT 1587
 1849 CCGATTAACCTGT 1908
 1588 CACGAGATGT 1647
 1909 AGGCACTGT 1968
 1648 GCGAGCTGT 1704
 1969 GCGAGCTGT 2028
 1705 GAACCTGT 1764
 2029 TGAGT 2088
 1765 GAATGT 1824
 2089 TGT 2148
 1825 CAGT 1884
 2149 CAATGT 2208
 1885 CCGAGCTGT 1938
 2209 CGAATGT 2268
 1939 CTTGT 1998
 2269 CACGAAAG 2338
 1999 CTTGT 2058
 2339 TTTGT 2388
 2059 AGGAG 2118
 2389 ATAG 2448
 2119 TGT 2178

2449 TGT 2508
 2179 TTTGT 2238
 2509 GTCAAGATGT 2568
 2239 ACTGT 2298
 2569 TTTGT 2628
 2299 ATTGT 2346
 2629 GTTGT 2688
 2347 GAGGAG 2406
 2689 AAGATGT 2748
 2407 ATTCGT 2463
 2749 TTTGT 2808
 2464 GT 2473
 2809 GT 2818

RESULT 15
 AF022947
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Gallus gallus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 2615)
 Feiner, L., Koppel, A.M., Kobayashi, H. and Raper, J.A.
 Secreted Chick Semaphorins Bind Recombinant Neutrophil with Similar
 Affinities but Bind Different Subsets of Neurons in Situ
 Affinities but Bind Different Subsets of Neurons in Situ
 Neuron 19 (1997) In press
 2 (bases 1 to 2615)
 Feiner, L., Koppel, A.M., Kobayashi, H. and Raper, J.A.
 Direct Submersion
 Submitted (05-SEP-1997) Neuroscience, Univ. of Pennsylvania, 105
 Johnson Pavilion 36th and Hamilton, Philadelphia, PA 19104, USA

FEATURES
 source
 CDS

1.2615
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 22..2379
 /product="collapsin 5"
 /codon_start=1
 /protein_id="AAB80952.1"
 /db_xref="GI:2522205"
 /translation="MIGRMSASADLLIALGLLELPAGYHATDTPRLRLSHKEL
 MDLNTSVSHSPFGLGLHMLIDEPYRPLVGGDLILSLDRI SNKYREIHSPST
 PLQAEELIKGDADECAHYKVLHRYRTHLLACGSAFPCVFIRVGHPSDHLF
 QLESHKFERGRGRCPEPDSFTSLILGELFTGLISYNGMRDAVAPRTMRMLRT
 EDSHKLKEPFTVGSYMI PDNEHDHDKVYLFTERKLEDETHALITPVGVGVN
 DMGGRIYVKNSTFLKRLVCSYVGRNGIDHDELDDELVELLOTDRKNVITGLS
 TTSNFRGVALICVYHAIYVRAFNCPYAHKEDEPEYVMAIYSGKVPYPRGSCASVNG
 GIKTTTKYDPDAVHAPRSHPLMOPIDIPVHRPILIVYEDKYNIKOIAVDVREDEG
 QYDVLEIAGVRFHOCMDTGTACADCCALAPYCANMGISGSRYPYTGQARRREROD
 IGRVNAADQCFOQFTGEVLETERLRYGLIENYNTLEXTPTLOAKRVNVEGAEH
 VRHGNAAQCFOQFTGEVLETERLRYGLIENYNTLEXTPTLOAKRVNVEGAEH
 TKKEVYKDEIRIKMDGLFLKRLHLDLGTIFCQTVERSIYHVRKLTLEIEERY
 DEMFSKDYEEELSHKMPDMOSNIDPVSKPWPKEFLQILIGSNFVREVEYCEKWCID
 KKKRKLKMSPSKWKYANPQEKRODEKKRIRPHEHYRILPRNIADS"

BASE COUNT 802 a 518 c 620 g 675 t
 ORIGIN

Query Match 21.6% Score 566.8; DB 5; Length 2615;
 Best Local Similarity 56.6% Pred. No. 3.5e-95;
 Matches 1208; Conservative 1; Mismatches 888; Indels 39; Gaps 7;

362 CCAATGTCCTCCGCTGCTCTCTACCGAGACCTCTGTCGCCAACCGCTGCGCA 421
 116 CGAGACACGCAAGGTTACGCTGTCTACATAAAGCTCTGGGATTTAAACAGAGCTGAG 175
 422 TCTTTGGGCCCCCAGGCTCCCTGAACCTCAGGCGCATGTACATAGATGATGACGAG 481
 176 TCTTTACATACCCATTTGGGTATCTTGCTCATATATAGCTCTGATGATATCAG 235
 482 ACCGCTCTTTCTGGGTGCTGAGCCCTTACCTCTGCTGCGGCGGAGACGACGCTGCG 541
 236 AGGACATGCTGCTGGGAGAGAGACCTCTGTACTCCCTCATTGGATGCAATCAGCA 295
 542 CAAATCCCGGAGGATCTCTGCTGACCCGACGAGAGAGAGAGAGATGTTGCAAG 601
 296 ACACTATACAGAGATTCATTGCTGCTAGTACACCTCTTCAAGAGCAAGATGATATTA 355
 602 AGGAGAGATCTTTGACAGAGTGGCCAACTTGTGGGGGTGTACAGCTCAGAC 661
 356 AAGGAGAGATGCT--GATGATGTCTAATTAATGTCCGTCTCCCTCAGCATCAATA 412
 662 GGAACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
 413 GAACGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 472
 722 TTGGCCACCTGCT--GAGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
 473 TTGGACATTCATCAGAGATCATCTGTTTCACTGATGATGATGATGATGATGATGATG 532
 779 GGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
 533 GAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 592
 839 TGTACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
 593 TTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652
 899 GTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955
 653 ATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712
 956 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
 713 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772
 1016 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1075
 773 TTTTACTGAAAAAGCATTTAGAGGCTGAGACAAAGC--ACTCATGCCATTTACACCGAG 829
 1076 TGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1135
 830 TGGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189
 1136 CTTTCTCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1195
 890 CTTTCTCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 949
 1196 TTACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1255
 950 TTGATGATTTAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1009
 1256 AGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1315
 1010 TTGGCTCTCTACAGCAATATATTAGAGATATGCTATATGCTGCTGCTGCTGCTGCTGCT 1069
 1316 TGGCAGACATTTGGGAGGTTTCAAGGCGCTTGGCCACCGAGATGGGCTCAGCAGC 1375

1070 TGGCAATCGTCCGAGCAGCTTTCATGACCAATGCTCATTAAGAGAGACCGAATACT 1129
 1376 AGTGGGGGCTTATGAGGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1435
 1130 ACTGGCTCTGACAGAGAGAGAGAGGATCTTATCTAGGCTGCTGCTGCTGCTGCTGCTG 1189
 1436 TGAACGACAG 1495
 1190 TGA-----ATGGCGTCTGTACTACTACCAACCAAGCTGCTGCTGCTGCTGCTGCTGCT 1240
 1496 AGTTTGGCCGAG 1555
 1241 ATTTTGGCAG 1300
 1556 TCTTGTGCAAG 1615
 1301 TTCTAGTAAAAAGATAG 1360
 1616 CAG 1675
 1361 CTGAAG 1420
 1676 TCAATGCTCTCCAGGAG 1735
 1421 TCATTACATTTTACATTCAG 1480
 1736 AGTGTATTAG 1795
 1481 AAGTATTCAG 1540
 1796 TATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1855
 1541 TCTCATTTGAG 1600
 1856 GCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1915
 1601 GCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1660
 1916 CTTGTACCACTACAG 1969
 1661 CTTGTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1720
 1970 TCCGCGAG 2029
 1721 TACGACATGGAAG 1780
 2030 GACTTGTGAG 2089
 1781 AGAAG 1840
 2090 TGGCAG 2149
 1841 CCCCTGAG 1900
 2150 CTGACAG 2209
 1901 AGAAG 1960
 2210 GCTTACAG 2269
 1961 ACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2020
 2270 ACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2329
 2021 ACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2080
 2330 CTGCGAG 2382
 2081 GTAAG 2140
 2383 -----CGAG 2437
 2141 CTCAGATTCAG 2200

Thu Oct 10 09:34:59 2002

us-09-813-290-1.rge

Page 23

QY	2438	CCCCGGTGA	TGAGTACTGTGAGCCG	TGTGTCA	2473
Db	2201	ACAGAGTGA	GAGGAGTACTGTGAAAAAG	TCTGTGTGA	2236

Search completed: October 9, 2002, 12:40:47
Job time : 3291 secs


```

OY 1621 GATGGACCTAGAGATCTATTTCTGGGAGACTGACTGAGGCTGTCTCAAGTCATC 1680
DB 1621 GATGGACCTAGAGATCTATTTCTGGGAGACTGACTGAGGCTGTCTCAAGTCATC 1680
OY 1681 GCTCTCAGGAGGAGGCTCAGCTGAGTGAAGAGAGGCTTCTGAGAGAGCTCCAGGTG 1740
DB 1681 GCTCTCAGGAGGAGGCTCAGCTGAGTGAAGAGAGGCTTCTGAGAGAGCTCCAGGTG 1740
OY 1741 TTTAAGTGGCCAAACCTATTCACGAATGAGATCTGTCTCAAAAGCAAAATGCTATAC 1800
DB 1741 TTTAAGTGGCCAAACCTATTCACGAATGAGATCTGTCTCAAAAGCAAAATGCTATAC 1800
OY 1801 GTGGGCTCTGCGCTGGGTGTGTGGCCAGCTGCGGCTGACCAATGTGAGACTTACGAGCT 1860
DB 1801 GTGGGCTCTGCGCTGGGTGTGTGGCCAGCTGCGGCTGACCAATGTGAGACTTACGAGCT 1860
OY 1861 GCTCTGAGAGATGCTGTGCTGGCCGAGCCATATCTGTCTGGATGTGTGCTCTCTGT 1920
DB 1861 GCTCTGAGAGATGCTGTGCTGGCCGAGCCATATCTGTCTGGATGTGTGCTCTCTGT 1920
OY 1921 ACCCACTACCGCCGAGCTTGGCAAGGCGGCTTCCGCGAGGACATCCGCGACGCG 1980
DB 1921 ACCCACTACCGCCGAGCTTGGCAAGGCGGCTTCCGCGAGGACATCCGCGACGCG 1980
OY 1981 AACCTGCGCTGAGTGTGGGCGGAGCCAGAGCAAGAGAGGAGGAGGAGCTTGTGCA 2040
DB 1981 AACCTGCGCTGAGTGTGGGCGGAGCCAGAGCAAGAGAGGAGGAGGAGCTTGTGCA 2040
OY 2041 GCCACATGTGTACGAGGAGGAGCAATAGACCTTCTGAGAGTGTGCTGCGCAAGTCT 2100
DB 2041 GCCACATGTGTACGAGGAGGAGCAATAGACCTTCTGAGAGTGTGCTGCGCAAGTCT 2100
OY 2101 CCCGAGTGTGTGCTGCTGCTTCTGAGAGGCTGAGAGGAGGAGGAGGAGGAGGAGT 2160
DB 2101 CCCGAGTGTGTGCTGCTGCTTCTGAGAGGCTGAGAGGAGGAGGAGGAGGAGGAGT 2160
OY 2161 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
DB 2161 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
OY 2221 TTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
DB 2221 TTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
OY 2281 CGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
DB 2281 CGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
OY 2341 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
DB 2341 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
OY 2401 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2460
DB 2401 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2460
OY 2461 CGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
DB 2461 CGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
OY 2521 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2580
DB 2521 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2580
OY 2581 GTGCAATGCGAGCAATGCGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640
DB 2581 GTGCAATGCGAGCAATGCGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640

```

RESULT 2
AAH47792
ID AAH47792 standard; DNA: 3568 BP.
XX

```

AC AAH47792;
XX 07-JAN-2002 (first entry)
XX Novel human protein (NHP) polynucleotide sequence.
XX NHP: novel human protein; secreted protein; semaphorin; oxytocin;
XX neurohypophyseal; neuropeptide; gene therapy; drug screening; ds.
XX Homo sapiens.
XX WO2001/0806-A2.
XX 27-SEP-2001.
XX 20-MAR-2001; 2001WO-0508834.
XX 20-MAR-2000; 2000US-190638P.
XX 22-MAR-2000; 2000US-191188P.
XX 31-MAR-2000; 2000US-193639P.
XX (LEXI-) LEXICON GENETICS INC.
XX Walke DW, Wilganowski NL, Turner CA, Halbur E, Wang X, Donoho G;
XX Scoville J;
XX WPI; 2001-611483/70.
XX New polynucleotides encoding human proteins that share structural
XX similarity with semaphorin proteins, protein hormones of
XX neurohypophyseal family for drug screening, diagnosis and therapy of
XX biological disorders.
XX Disclosure; Page 40-41; 43pp; English.
XX The invention relates to novel human secreted proteins (NHP) that share
XX structural similarity with semaphorin proteins, protein hormones
XX of the neurohypophyseal family and oxytocin (neurophysin 1 precursor)
XX family. The NHP polynucleotide sequences are useful in drug screening
XX techniques for creating symptomatic or phenotypic manifestations of
XX perturbing the normal function of NHP in the body. Nucleotide constructs
XX encoding NHP products are useful in gene therapy for modulating NHP
XX expression. The constructs can be used to genetically engineer host cells
XX to express NHP products in vivo, these genetically engineered cells
XX function as bioreactors in the body delivering a continuous supply of a
XX NHP, a NHP peptide, or a NHP fusion protein to the body. The sequences
XX also find use in molecular multigenesis/evolution of proteins that are
XX partially encoded by the NHP sequences. The encoded NHP polypeptides are
XX useful for generating antibodies, as reagents in diagnostic assays, for
XX identifying other cellular gene products related to NHP and as reagents
XX in assays for screening for compounds that are useful in the treatment of
XX mental, biological or medical disorders and diseases. The present
XX sequence represents a NHP polynucleotide sequence.
XX Sequence 3568 BP; 633 A; 1099 C; 1156 G; 680 T; 0 other;
SQ
XX Query Match 100.0%; Score 2627.6; DB 22; Length 3568;
XX Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
XX Matches 2627; Conservative 1; Mismatches 0;
OY 1 ATGGCTGTGCTTGTGCTGAGGAGGCTTCCCAATGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
DB 555 ATGGCTGTGCTTGTGCTGAGGAGGCTTCCCAATGAGGAGGAGGAGGAGGAGGAGGAGGAG 614
OY 61 AGCTGCACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 674
DB 615 AGCTGCACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 674
OY 121 CTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 675 CTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
OY 181 TGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
DB 181 TGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240

```


The sequence was isolated

CCCTGACCGCGGTGGGCCGCTCTGCGTGAAATGATGCT

Disclosure; Page 37-38; 43pp; English.

xx Sequence 2349 BP; 406 A; 747 C; 753 G; 442 T; 1 other
sq

Case	Score	DB	Match
89.48;	2348.6;	22,	0
100.0%;	Pred. No: 0;		

280 ATGGCCCCCTCGGAGCCATTTGTCTGGCAGCAAGGAGGCCACGCAGTTC
60

ATGAGTGGCCCGCTCCTACCGAGACCTC 39

61 AGCTCTGGCCCCAGCCCCGCCCCCAGTATACCCCGGC-
TCATGCGGTCCCATCAACTTGACGCC 45

121
CTCTTGGCCACCGCTCTGCCATCTTCTGGGCCCCAGGGCTCCCTGAACCTCCAGGC 128

460 ATGTACCTAGATGACATCCGCGCGCTGGACGCCCTCTACTCT 24

520 CTGGCGCTGGACCAAGCATGCGCCAGATCCCGGAGGTCTCTGTGGCCACCACCGACCGC

241 CTGCGGCACGGAGATCCTTTCACAGAGTGCGCCACTTCGTG 6

301 CAGAGGAGGAGTGTCTCGAAAGGGAAGAGAATCCCTTTGACAGAGCACCCTCTCT

070
CCCTCACCACGACCGACCTGTGGCACTGGGCCCTTCAG
|||||
|||||

4

700 CCCACCTGTGCCCTCATCACAGTTGCGCACCAATGGCCGACTTTC

740. CCGCAGTGTGGAAAGTGGCCCGGGGGGGTGGCCCTACAGAGCCAGCCGCTCCCTTTGCCAGC

481 GGCAGTGTGGAAATGCCCCCCCCCCCC

541 ACCTTCATAGACGGGAGCTGTACACGGGCTCTCACTGCTGACTTCCCTGGGGCGAGAGGCC

880 AIGALCITCCGAGCGGGG

y

OY	2470	TGCA	2473
Db	2809	TGTA	2812
XX	RESULT	9	
XX	ID	AAZ28470	
XX	AC	AAZ28470 standard; DNA: 3988 BP.	
XX	DT	05-JAN-2000 (first entry)	
XX	DE	Mouse semaphorin H variant (Sema HV) polynucleotide sequence.	
XX	KW	Semaphorin H; Sema H; Sema HV; collapsin; bone structure formation;	
XX	OS	metastasis; cancer; antibody; drug screen; ss.	
XX	Mus sp.		
FH	Key	Location/Qualifiers	
FT	CDS	609..2943	
FT		/tag=a	
FT		/product= Sema_Hv	
PN		/note= "Mouse semaphorin H variant"	
XX	MO9947671-AZ.		
PD	23-SEP-1999.		
XX	PF	12-MAR-1999; 99MO-IB00495.	
PR	13-MAR-1998; 98US-0077997.		
XX	(LUKA/) LUKANIDIN E M.		
PA	(CHRI/) CHRISTENSEN C R L.		
XX	Lukanidin EM, Christensen CRL;		
DR	WPI: 1999-590975/50.		
P-PSDB:	AAV43091.		
XX	New polypeptides and polynucleotides, useful in diagnosis and treatment		
PT	of metastatic cancer -		
XX	Claim 1; Page 85-91, 95pp; English.		
XX	This is the mouse Semaphorin-H variant (Sema HV) polynucleotide sequence		
CC	(Sema Hv). The Semaphorin/Collapsin family of molecules are characterised		
CC	by a unique and highly conserved motif, within a 500 amino acid		
CC	semaphorin domain. Some semaphorins exhibit inhibitory or repulsive		
CC	functions in a neuronal context, and functions in bone structure		
CC	formation are also implicated. It is thought that Sema H plays a role in		
CC	metastasis. The invention uses the mouse Sema H gene in the diagnosis of		
CC	metastatic cancer. Semaphorin polynucleotides and polypeptides are used		
CC	in the methods of the invention, the polypeptides are useful for		
CC	determining the metastatic potential of cells, by detecting their		
CC	expression in biological samples. Antibodies specific for Sema H, are		
CC	also useful therapeutically in inhibiting Sema-H polypeptide activity and		
CC	therefore metastasis, and for purifying the polypeptides. Metastasis may		
CC	also be inhibited by inhibiting the biological activity of the		
CC	ligand (or fragment). The polynucleotides can also be used to inhibit		
CC	polypeptide expression in cells using known antisense technology e.g. to		
CC	prevent metastasis of cancer cells. They can be used to detect and		
CC	quantify Sema-H mRNA levels in cells. The polypeptides, fusion proteins,		
CC	multimeric proteins, antibodies or antisense oligonucleotides can be		
CC	included in pharmaceutical compositions. The polynucleotides can be		
CC	used to isolate similar sequences from other species and to produce mammalian		
CC	cell lines and tumours with known metastatic potential, useful in		
XX	anti-metastatic drug screening.		

[illegible]

QY 2347 GAGGAGCCCCGAGCCGCGGAGCTGCTCCACCCCAAGCCCTGATCAAGAC 2406
 Db 2689 AAGATGCCCCGCTCCCTCCCTTAAAGCGATATGCTCAGAGGAGCAAAACCGGTACAGAGAA 2748
 QY 2407 AATCGACACCTGATGCTGCTCCCAACCTGCCCC---GGGTGATGAGATGCTGAGGCG 2463
 Db 2749 TTCTTGCACCTGATGCTGCTCAGCAGCAAGTTCAGAGAGTGAAGATCTGCGAAG 2808
 QY 2464 GTGTGCTCA 2473
 Db 2809 GTGTGCTCA 2818

RESULT 10
 AAX75767
 ID AAX75767 standard; DNA: 2530 BP.
 XX AAX75767;
 XX 22-JUL-1999 (first entry)
 DE Human semaphorin III DNA.
 XX Human: beta-amyloid precursor protein; beta-APP; diagnosis: cancer;
 KW frameshift mutation; age-related disease; neurodegenerative disorder;
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 KW neurofilament-F; presentin I; presentin II; cellular tumour antigen-1;
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMP-C; NSP-A;
 KW high mobility group protein-C; neuroendocrine specific protein A; ss.
 XX Homo sapiens.
 OS

XX Homo sapiens.
 XX MO9845322-A2.
 XX 15-OCT-1998.
 XX 02-APR-1998; 98MO-IB00705.
 XX 10-APR-1997; 97US-0043163.
 XX (UYU-); RIJSUNIV UTRECHT.
 XX (ROYA-); ROYAL NETHERLANDS ACADEMIE ARTS & SCI.
 XX (UYRO-); UNIV ROTTERDAM ERASMUS.
 XX Burbach JPH, Grosveld FG, Van Leeuwen FW;
 DR MPI; 1998-609901/51.
 XX

Diagnosing disease by detecting frameshift mutations in RNA or
 corresponding protein mutations - used to diagnose cancer and
 neurological diseases, particularly Alzheimer's disease, and also
 for treatment and prevention with specific ribozymes or wild-type
 RNA

Disclosure; Figure 16; 258pp; English.

This invention describes a novel method for the diagnosis of a disease
 caused by, or associated with, an RNA molecule that has a frameshift
 mutation. The method is used to diagnose age-related diseases,
 especially cancer and a wide range of neurodegenerative disorders,
 Alzheimer's disease, Down's syndrome, myotonic dystrophy, Huntington's
 disease, multiple sclerosis, alcoholic liver disease, diabetes mellitus
 type II and many others listed) or susceptibility to these disorders. The
 method allows a definitive diagnosis of Alzheimer's disease in living
 patients, at an early stage. It is based on the observation that disease
 may be caused by mutations in RNA rather than DNA. The invention
 describes the use of neuronal system RNA molecules, specifically
 proteins including beta-amyloid precursor protein (beta-APP), the
 microtubule associated proteins Tau and Big Tau, ubiquitin B,

apolipoprotein E, microtubule associated protein 2 (MAP2),
 neurofilament-L, neurofilament-M, neurofilament-F, presentin I,
 presentin II, glial fibrillary acidic protein (GFAP), the cellular
 tumour antigen p53, B-cell leukemia/lymphoma 2 (bcl-2) proto-oncogene,
 semaphorin III, HMP-C, high mobility group protein-C (HMP-C) and
 neuroendocrine specific protein A. This sequence encodes the wild type
 and mutant protein fragments represented in AAX71264-121348.
 SO Sequence 2530 BP; 786 A; 518 C; 576 G; 650 T; 0 other;

Query Match 18.1%; Score 474.8; DB 19; Length 2530;
 Best Local Similarity 53.9%; Pred. No. 8.3e-95;
 Matches 1146; Conservative 1; Mismatches 948; Indels 33; Gaps 7;

QY 363 CAGTGGCCCCGCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 422
 Db 96 CAGTGGCCCCGCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 422
 QY 423 CTCTTGGCCCCGCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 155
 Db 156 TTTCATGCGCTTGGCCACACCTCCAGCTTATCATACCTCTTGGATGAGAACGAG 215
 QY 483 CCGCTCTTGGGCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 215
 Db 216 TAGCTGATGCTGAGAGCAAGATCATATTTTCATGACCTGTTATATCAAG-- 273
 QY 543 AGATCCCGGAGGCTCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 602
 Db 274 -GAATTTCAAAAGATGCTGCGGCTCTCTACAGAGCCCTGCTGCGACACCCCTGCGCAT 602
 QY 603 GGAAGAGATCTTGGACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 662
 Db 333 TGGAAAGATCTTGGACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 662
 QY 663 GACCCACTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 722
 Db 393 GACTCACTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 722
 QY 723 TGGCCACCG---TGGGACATGCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 452
 Db 453 TGGACATGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 452
 QY 780 GGGGGGCTCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 839
 Db 513 TGGGAGATCTTGGACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 839
 QY 840 GTACAGGCTCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 572
 Db 573 ATACTGGAAGCTGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 572
 QY 900 TCGTGGCCACTGCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 632
 Db 633 GCACCACTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 632
 QY 957 TGTATGCGCGCCGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 1016
 Db 693 CATTAGGCTGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 1016
 QY 1017 CTCTCGAGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 752
 Db 753 CTCTCGAGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 752
 QY 1077 GGGCGCGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 809
 Db 810 AGGTCAAGATGCAAGATGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 809
 QY 1137 TTTCCAGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 869
 Db 870 ATTCCTCAAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 869
 QY 1197 TGACAGCTGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 929
 Db 930 TGATGAGCTGAGCTGCTGAGCTGCGGCTCTCTACAGAGCCCTGCTGCGGCTCTCTACAGAGCCCTGCTGCGCAT 929

[illegible][illegible]

Matches 1146; Conservative 1; Mismatches 948; Indels 33; Gaps 7;			
OY	363	CAGTGTCCCGCCCTGCGGCTCCCTACCGAGACCTGCTGTGCGAACCCCTTGCCAT	422
Db	96	CAATGTGCCAAGGCTGAATTTATCTCAAGAAATGTGAAATCCAAATGTGAT	155
OY	423	CTTTGGGCCCCCGAGGCTCCCTGAAACCTTCAGGCCATGACTAGATAGTACCAGA	482
Db	156	TTTCAATGCTTGGCCCAACAGCTCCAGTTATCATACCTTCTTTTGGATAGGAAGGAG	215
OY	483	CCGCTCTTTTGGGTGGCTGGACGCCCTTACTCTCTGCGGCTGGACGAGCATGGC	542
Db	216	TAGGTGTATGTGAGCAAGGATACATATTTTTCATTCGACCTGGTAAATATCAAC	273
OY	543	AGATCCCGGAGGTCTCTGGGACCCGACCCAGCAGAGAGAGAGAGAGAGAGAGAG	602
Db	274	-GATTTTCAAAAGATTGTGGCAATCTTATCAACAGGATTAAGGATATATCA	392
OY	603	GAGAGAGATCTTTGACAGAGTGGCCCAACTGCTGGGGTGGTACAGCTCAACACG	662
Db	333	TGGAAAGACATCTGAAAGATGTCTAATTTATCATAGTACTTAAGCATATATCA	392
OY	663	GACCCACCTGCTAGCTCTGGACCTGGGCGCTTCACGCCACCTGTCCTATCAGCT	722
Db	393	GACTGACTTGTAGCTGCTGGAGCGGGGCTTTTCATCCAAATTTGCACTTCAAT	452
OY	723	TGGCCACCG---TGGGAGCATGTGCTCCACCTGGAGCCTGGAGTGGGAAAGTGGC	779
Db	453	TGGACATCATCTGAGGACATATTTTAAAGCTGGAGAACTCACATTTTGAAGAGCG	512
OY	780	GGGGAGGTGCTTCAGAGAGCCAGCCGCTTTCAGACACTTCATAGAGGGAGCT	839
Db	513	TGGGAGAGTCCATATGACCTTAAGCTGCTGACAGCATCTTTTAAATGTGAGAA	572
OY	840	GTACAGGCTGCTACTGCTGACTCTGAGCTCTGGGGAGAGGCCATGATCTTCGAA	899
Db	573	ATATCTGTGAACTGAGCTGATTTTATGGGCGAGACTTGTATCTCGAAGCTTGG	632
OY	900	TCTTGGGCAAGCTGCTGCTGCTGAC---TCTGACCAAGTCTTCTGACAGACCCCG	856
Db	633	GCACCAACCACTTACAGAGAGACAGCATGATCTCAGGTGCTCAATGATCCAA	692
OY	957	TGTATGGCCCGCCGAGTCTGAGAACTGTGACAGCAATGCAAGGTACTTCTT	1016
Db	693	CATTAGTCCACCTCATCTGACAGAGTACAACTCTGAAATGACAAAGTATCTT	752
OY	1017	CTTCTCGAGAGAGTCTCCCTGCGCCGATGCTGCTGCAACCATGCTGCTGCA	809
Db	753	CTTCCGGAATG---CAATAGATGGAACACCTGGAAGCTACTCACTGCA	869
OY	1077	GAGCGCGCTGCTGCTGATGATGCTGGGAGCCAGCGGCTGCTGCTGCAAAATG	1136
Db	810	AGGTGAGATGCAAGAAATGACTTGGAGGACACAGAACTGCTGTAATTAATG	869
OY	1137	TTTCTCAAGCCAGGCTGCTGCTGCTGGGCGCCGCTGCTGCTGCTGCTGCTG	1196
Db	870	ATTCCTCAAACTGCTGCTGATTTCTGAGTCCAGCTCAAAATGATGACACT	929
OY	1197	TGACACAGTGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1256
Db	930	TGATGAACTGCAAGATGATCTCTATGAATTTAAATCTTAAATTTCAAGTATA	989
OY	1257	CGGCGTGTGACAGCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1316
Db	990	TGAGAGTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1049
OY	1317	GCGACATCTGAGAGTGTTCACAGGCGCTTTGCCCAGAGATGGGCTGAGACCA	1376
Db	1050	GAGTATGTGAGAGAGGCTTCTTGTGCTCAATATGCCCACAGGATGAGACCA	1109
OY	1377	GTTGGGCGCTTATGAGAGAGTCCCTTCCCTGCGCCGCTGCTGCTGCTGCTGCT	1436
Db	1110	ATGGTGTCTTATGAGAGAGTCCCTATCCAGGCGCAGAACTGCTCCAGCAAA	1168
OY	1437	GACCGACACCGCAGAGAGGCTTTTGGACACCAAGACTACCCAGATGAGTGTGCA	1496
Db	1169	-----CATTTGGGTGTTTGTACTCTTACAGAGAGCTTCTCTGATGTATAC	1217
OY	1497	GTTTCCCGAGGCGCCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1556
Db	1218	CTTTGCAAGAGTCAATCCAGGCTATGATGATGATGATGATGATGATGATGAT	1277
OY	1557	CCTTGCAAGAGGCGCCCGCCAGCAGCTACACAGATGCTGCTGCTGCTGCTGCT	1616
Db	1278	AGTATCAAAAGGATGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	1337
OY	1617	AGAGATGAGACCTACAGATGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1676
Db	1338	AGAGATGAGACCTACAGATGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1397
OY	1677	CATGCTCTCCAGGCGAGGCGCTCAGCTAACCCTGAGAGAGTGTCTGCTGCTGCT	1736
Db	1398	AGTTTCAATTTCTTAAGAGAGCTTGTATGATTTTAAAGAGGCTTCTGCTGCTGCT	1457
OY	1737	GATTTTAAAGTGTGCAACACCTATATCAAGAAATGAGATCTGCTGCTGCTGCTGCT	1796
Db	1458	AGTTTTCGGAACGACCTGCTATTTTCAAGCAATGAGCTTTCCACTTAAAGCA	1517
OY	1797	ATAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1856
Db	1518	ATATATTTGCTTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1577
OY	1857	CAGTCTGTGCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1916
Db	1578	GAAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1637
OY	1917	CTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1976
Db	1638	ATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1694
OY	1977	CGGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2036
Db	1695	TGAGAGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1754
OY	2037	GCGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2096
Db	1755	TGAGAGAGATTAATCTATGATGATGATGATGATGATGATGATGATGATGATGAT	1814
OY	2097	GCTTCCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2156
Db	1815	GTCGCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1874
OY	2157	GCTGCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2216
Db	1875	GATCAAGAGTGAATGATCATATCATGAGACAGATCAAGGCTTGTGCTGCTGCTGCTGCT	1934
OY	2217	CCGTTTGAATGGGCGCTTACACTGACACCTGACACCTGACATGCTTCTCCAGACTG	2276
Db	1935	ACAGAGAGATTTAGCAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1994
OY	2277	GCTTCCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2054
Db	1995	TCTTAAAGTAACTCTGGAATCATTTGACACAGAGCATTTTGAAGAACTTCTTCA	2114
OY	2337	GCGAAGAGAGAGAGGCGCCAGCGCGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT	2174
Db	2055	TGATGATGAGAGTGTCTTAAGACCAAGAAATGTCCATATGACAGACCTTGA	2202
OY	2391	GCGCTGTACAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2240
Db	2115	GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2280
OY	2451	GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2320
Db	2175	GTTTGTGAGCAAGTTTGAAGAGGAC	2360

```

XX RESULT 12
XX AAH47049
XX ID AAH47049 standard; DNA; 2709 BP.
XX AC AAH47049;
XX DT 29-OCT-2001 (first entry)
XX XX Semaphorin D cDNA sequence.
XX XX Drug resistance; resistance gene; semaphorin D; B94; mel-14 antigen;
XX XX 24p3; proliferin; maspin; cancer; cytostatic; gene therapy; ss.
XX OS Homo sapiens.
XX PN W0200155455-A2.
XX PD 02-AUG-2001.
XX XX 31-JAN-2001; 2001WO-0503161.
XX PF 31-JAN-2000; 2000US-0179191.
XX PR 31-JAN-2000; 2000US-0179191.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PJ (JINS/) JIN S.
XX PI Jin S;
XX DR WPI: 2001-488799/53.
XX PT Determining if a compound modulates the drug resistance of a cell,
XX PT comprises determining the expression or activity level of a resistance
XX PT sequence in a cell in the presence of the test compound
XX XX Example 1; Fig 1A-B; 79pp; English.
XX XX The invention relates to a method of determining whether a test compound
XX XX modulates the drug resistance of a cell that comprises determining the
XX XX expression or activity level of resistance genes (e.g. semaphorin D, B94,
XX XX mel-14 antigen, 24p3, proliferin or maspin) in a cell in the presence of
XX XX the test compound, and comparing its expression or activity level in a
XX XX cell without the test compound. The drug resistant sequences are useful
XX XX in identifying drug resistant cells, in screening methods directed to the
XX XX identification of compounds that can modulate the drug resistance of a
XX XX cell type or multiple cell types. An isolated resistance protein can be
XX XX used as an immunogen to generate antibodies that bind the resistance
XX XX protein. Resistance nucleic acids may be inserted into vectors and used
XX XX as gene therapy vectors. An anti-resistance protein antibody may be used
XX XX to isolate a resistance protein from cells and/or facilitate the purification of
XX XX natural resistance protein expressed in host cells. The methods are useful for
XX XX resistance protein expressed in host cells. The methods are useful for
XX XX treating a subject having a disorder, such as a drug-resistance cancer,
XX XX characterized by aberrant resistance sequence expression or activity by
XX XX administering to the subject a resistance modulator. The present sequence
XX XX represents a semaphorin cDNA sequence, whose expression was increased in
XX XX drug resistant EMT6 tumours.
XX XX
XX XX Sequence 2709 BP: 831 A; 560 C; 620 G; 698 T; 0 other;
XX XX
XX Query Match 18.1%; Score 474.8; DB 22; Length 2709;
XX Best local similarity 53.9%; Pred. No. 8.4e-95;
XX Matches 1146; Conservative 1; Mismatches 948; Indels 33; Gaps
XX
XX 363 CATGTGCCCCCGCGGCTCTCTACCGAGACCTCTGTCTGCCAACCGCTCTGCCAT 422
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 280 CAATGTGCGCAAGGCTCAATTAATTCCTACCAAGAAATGTTGCAATCCAAATGTGATC 339
XX
XX 423 CTTTCTGCGCCCCCAGGCTCCTGTACACTTCAGGCGCATGTACCTAGTACGATCCGAGA 482
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 340 TTTTCATATGCGCTTGCGCAACGCTCCAGTTATATCATATCTCTTTTGATGAGGAACGAG 399
XX
XX 483 CCGGCTCTTTCTGGTGCGCTGTGACGCCCTCTACTCTCTCTGCGGCTGAGCAACGATGGCC 542

```

Db	400	TAGCTGTATGTTGGAGCAAGATACATATTTTTCATTGCACTGGTATATATCAAG--	457
Qy	543	AGATCCCCGGGAGAGTCTCTGTGGCCACCGACCGACAGAGAGGAGAGTGTTCGAA	602
Db	458	-GATTTTCAAAAGATTGTGTGGCCATATCTTACACAGAAAGATGTATTCAGTGGCC	516
Qy	603	GGGAGAGATCTTTGACAGAGTGGCCCAACTGTGGGGGTCTACAGCTCAACACCG	662
Db	517	TGGAAACACATCTCTGAAGAAATGTCTAATTTTCATCAGAGTACTTAAGGATATATCA	576
Qy	663	GACCCACTGTAGCTGTGGCACTGGGGCTTCCAGCCACTGTGGCTCATACAGT	722
Db	577	GACTACTTGTAGCGCTGTGGAGAGGGGGCTTTTTCACCAATTTTGACACTCATATGAAT	636
Qy	723	TGGCCACG---TGGGAGGACATGTGTCCACCTGGAGCCCTGGAGTGTGAAAATGGCC	779
Db	637	TGGACATCACTCTGGAGCAATATTTTAAAGTGGAGAACTCATATTTGAAAGAGCGC	696
Qy	780	GGGGGCTGCTCCACGAGCCAGCCAGCCGTTCCTTTGCGAGCACTCATATGAGCGAGCT	839
Db	697	TGGGAGAGTCCATATGACCTTAACCTGTGACAGCATCCCTTTATATAGATGGAAAT	756
Qy	840	GTCACAGGCTCATCGCTGACTTCTGGGGCGAGAGCCATGATCTTCCAGTGGAG	899
Db	757	ATACTGTGAATGACAGCTGATTATTATGGGGCGAGACTTGTCTATCTTCGAACTCTTG	816
Qy	900	TCTCGCCGACCTCTGCTGTCCGAC---TCTGACCAAGCTCTTTGACACACCCCGGT	956
Db	817	GCACCAACACCCATAGAGACAGAGAGCATGATATTCAGGTGGCTCATATGACCAAGT	876
Qy	957	TGTGATGGCGCCGGATCCGTAGAACCTTGACAGGACATGACAGATGATGATCTTCT	1016
Db	877	CATTAGTGGCCACTCATCTCAGAGAGTACACATCTTCAAGATGCAAAAGATATCTTTT	936
Qy	1017	CTTCTCGAAGAGTCTCCCTCGCCCGATGTGGGTCTGCAACCATGTCACTGTACCCCGT	1076
Db	937	CTTCGTTAAATAT---CAATAGATGAGAGACACTGTGAAAGATGCTACTCAGCTAGAT	993
Qy	1077	GGCGCGGTCTGTGAATATGCTGGGGCCAGCGGTGTGTGTGAACAAATGGAGAC	1136
Db	994	AGGTAGATATGCAAGATATGACTTTGGAGGGCACAAGCTGTGGTAATTAATGGACAC	1053
Qy	1137	TTTCTCAAGGCGAGCTGTGTCTGTGGGTGGCGCCGTGGGTGGTGGACGCCACTT	1196
Db	1054	ATTCCTCAAAAGCTCTCTGATTTCTCAGTGGCCAGGTGCCAAATGGCATTTGACACATTT	1113
Qy	1197	TAGCCAGTAGAGATGTCTCTGTGTGGCCAGAGCGCGGAGAAAGCTGTGAGGTGA	1256
Db	1114	TGATGACTGCAAGATATTCATTAATTAATTAAGATCTTAATAATTCAGTTTATA	1173
Qy	1257	CAGCGTGTAGACCGTCAGTGCCTGTCCAGAGCGCTGCGCGTGTGTGTACCAT	1316
Db	1174	TGGAGTGTATACACTTCAGTAATTTTCAAAGGATCAGCCGTGTGTGTATAGCAT	1233
Qy	1317	GGGAGCATGTGGAGATTTTCAAGGGCCCTTTGGCCACCAGATATGGGCTCAGACCA	1376
Db	1234	GAGTGAATGTGAAGGCTGTCTTGTGTCAATATCCCAAGAGATGTAGACCAATATCA	1293
Qy	1437	GACCGCAGCAGCAGAGGCGCTTTGGCAGACCAAGACATACCAAGTATGGGTGTGCA	1496
Db	1353	-----CATTTGGTGTATTACATCTACAAAGACCTTCTGTATGATTTATAC	1401
Qy	1497	GTTTGCCGAGCCACCCCTCATGTTCTGGGCTGTGGCGGCTCGACATGGCGGCTGT	1556
Db	1402	CTTTTCAAGAGTATCAGGACATGTATCAATTCAGATGTTTCTATGACCAATTCGCAAT	1461
Qy	1557	CTTTTCAAGACCACTGTGGCCAGAGCTACACAGATCTGTGTGGACCGGTGTGAGGC	1616

	RESULT 13
AAK81571	
ID	AAK81571 standard; DNA; 4556 BP.
XX	
AC	AAK81571;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO: 36383
XX	

KW	Human; immune; hematopoietic; immune/haematopoietic antigen; cancer
KM	Cytostatic; gene therapy; vaccine; metastasis; ds.
KX	
XX	Homo sapiens.
OS	
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	11-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	14-JUL-2000; 2000US-0217496.
PR	26-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0218963.
PR	14-AUG-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229513.
PR	05-SEP-2000; 2000US-0229519.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0233080.
PR	08-SEP-2000; 2000US-0233081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	21-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234274.

XX	Jacobs K, McCoy JM, Lavallie EJ, Collins-Racie LA, Evans C:
PI	Merberg D, Treacy M, Spaulding V,
DR	WPI: 2000-679620/66.
XX	P-PDSB; AAB28379.
XX	New monoclonal antibodies, useful for treating cancer and immunodetection of secreted proteins which are in turn useful for treating neurological, inflammatory, immune diseases and microbial infections -
PS	Disclosure: pages 66-67; 75pp; English.
XX	The present sequence is the coding sequence for human clone BR53.4. The present sequence was isolated from a human foetal kidney cDNA library.
CC	The BR53_4 protein is a secreted protein and can be used to treat a number of conditions including various immune deficiencies and disorders, asthma, respiratory problems, inflammation, myeloid or lymphoid cell deficiencies, anaemias, burns, wound healing, ulcers, periodontal disease, central and peripheral nervous system diseases and neuropathies, lung or liver fibrosis and cancer.
CC	
CC	
XX	Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;
SQ	
Query Match	14.9%; Score 391.6; DB 21; Length 2975;
-Best Local Similarity	53.1%; Pred. No. 1.6e-76;
Matches 1004; Conservative	1; Mismatches 85; Indels 30; Gaps
OY	394 GACCTCCTGTCGCCAACCGCTCGCACTTCTTGGGCCCCCGAGGCTCCTGAACCTC 453
DB	347 GAACCTCGAANACCAGACTGTGAATCTCACCTTTCCACCACATCTTAGACTAC 406
OY	454 CAGGCACATGACTAGATGATGATACCGAGACCGCTTCTGGTGGCCCTGACGCCCTC 513
DB	407 AGGATTTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466
OY	514 TACTCTCTGGCGTGGGACCGACCATGCGCCAGATCCCGGAGAGTCTCTGCGCACCGCAG 573
DB	467 CTCTCCCTGAATATTAACTATATGATGATGATGATGATGATGATGATGATGATGAT 526
OY	574 CCAGACAGAGAGAGAGAGATGCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 633
DB	527 ACAATCAAAGTGTGAAGAATGCAAAATGCTGGCAAGATCCACACAGCGCTGGGAGAC 586
OY	634 TTCTGGGGGGTGTACAGCTCAGAACCGGACCCAGCTGTAGCCTGTGGAGATGGGGCC 693
DB	587 TTGTGCGTGTATTCACACTTTCACATGCGACACTTGTGTGTGTGTGTGTGTGTGTGT 646
OY	694 TTCAGACCCACCTGTGCCCTCATCACTGAGTGGGCCACCGGGSAGACATGTGCTCACCTG 753
DB	647 TTCAGTCTGCTGTACTACTTACTTGTGAACAAGAGAGAGAGAGATGACAGACCAATTTTCAATG 706
OY	754 GAGCCTGGAGTGTGAAGTGGCCGGGGGGGGGCTGCCCTCAGACAGCCCGACGGCTGCCCTT 813
DB	707 ATTGACTCCAAGTGTGAATCTGGAAGAAGAGACCTGCTCTTAAACCCACAGTGAACAG 766
OY	814 GCCAGACCTTCATAGACGGGAGAGCTGTACACGGGGCTCACTGCTGACTCTGCTGGGCGCA 873
DB	767 GTGTCTGTATATGATCAATGAGAGAGCTTTTCTCTGAGATGATATATGATTTTCAATGGGAGACA 826
OY	874 GAGGCACATGATCTTCGAAATGAGAGTCTTCGCGCCACAGCTGTGGCTTCCGACTCTGACAG 933
DB	827 GATGCTCTATTTTTTGAAGTTAAACCAAGAGAGANTGGGTCAAGATGATCAACATAT 886
OY	934 AGTCTTTC---ACGACCCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 990
DB	887 TCCAATAGGCTCAAGTGAACCTATGTTGTATGATGATGATGATGATGATGATGATGATGAT 946
OY	991 CAGGACATGACAGAGTGTACTTCTTCTTCGAGACAGGTCCCTCGCCCGATGAGTGGGC 1050
DB	947 CCAATGATGCTAAGGTGACTTCTTCTTCAAGAAAAATGACTGACATAATACAGGAGC 1006

Run on: October 9, 2002, 10:48:08 ; Search time 2035 Seconds
(without alignments)
17429.974 Million cell updates/sec

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Minimum	DB	seq	length:	0
Maximum	DB	seq	length:	20000000000

```
Post-processing: Minimum Match 0%
Maximum Match 100%
```

```

1:  Est.*
2:  em_estba.*
3:  em_estin.*
4:  em_estnu.*
5:  em_estov.*
6:  em_estro.*
7:  em_hc.*
8:  em_hc.*
9:  gb_est1.*
10: gb_est2.*
11: gb_hc.*
12: gb_gss.*
13: gb_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	466.8	17.8	3436	11	AK014333	AK014333 Mus musculus
c	418.4	15.9	420	9	AB080383	AB080383 i12-UM008
3	369.4	14.1	647	10	AB167436	AB167436 pflin.pk0
4	357.6	13.6	364	9	A1750874	A1750874 pnc6b10.x
5	320.2	12.2	576	10	B1057451	B1057451 pncf19.pk0
6	308.2	11.7	436	10	BM287610	BM287610 526378.MM
7	280.4	10.7	645	10	B1066459	B1066459 pflin.pk0
8	264.2	10.1	785	10	B1068389	B1068389 p6131583
9	255.2	9.7	609	10	B1008258	B1066237 pflin.pk0
10	250.2	9.5	990	10	B1557438	B155258 60329970
11	249.6	9.5	662	10	B1557438	B155738 60333896
12	243.8	9.3	920	9	AL541401	AL541401 AL541401
13	237.4	9.0	786	9	AL584324	AL584324 AL584324
14	235.2	8.9	754	9	AL584625	AL584625 AL584625
15	231.2	8.8	823	10	B1644606	B1645606 60327510
16	219	8.3	624	9	AM328571	AM328571 d803d12.x
17	217	8.3	735	9	AN140116	AN140116 AN140116

18	215	8.2	737	10	B1646227
19	212.6	8.1	343	10	H26438
20	210.2	8.0	626	9	AU137591
21	207.2	7.9	797	10	B1100405
22	205.6	7.8	748	9	AU136052
23	203.8	7.8	1121	10	BM477901
24	202.6	7.7	564	9	AM659026
25	202.6	7.7	678	10	B1151058
26	200.2	7.6	674	9	AU138132
27	197.8	7.5	650	9	AU134748
28	197.6	7.5	728	9	AU137554
29	193	7.3	629	9	B116052
30	192.6	7.3	778	10	BG184444
31	189.4	7.2	721	9	AM742133
32	188.2	7.2	433	10	BE009731
33	186.8	7.1	593	10	B1065528
34	186.6	7.1	637	9	AU155310
35	186	7.1	891	9	AL155137
36	183.6	7.0	9	BE017963	
37	185.4	6.9	1095	10	BM473502
38	180.2	6.9	755	9	AU137539
39	179	6.8	557	10	BF087910
40	178.2	6.8	494	9	AM957332
41	177.8	6.8	774	10	BG215143
42	174.8	6.7	434	12	AZ991314
43	169.4	6.4	468	10	BE710200
44	169.2	6.4	568	10	BF934558
45	169.2	6.4	602	10	BM425751

ALIGNMENTS

[illegible]

JOURNAL MEDLINE PUBMED REFERENCE	TITLE
Genome Res. 10 (11), 1757-1771 (2000)	FANTOM Consortium
11076861	Functional annotation of a full-length mouse cDNA collection
4 (sites)	(bases 1 to 3436)
The Riken Genome Exploration Research Group Phase II Team and the	Adachi, J., Aizawa, K., Akaboshi, S., et al.

Arakawa, T., Baldarelli, R., Canalis, S., Aikimura, T., Aono, H., Arai, A., Carninci, P., Fukuda, R., Brownstein, M., Buit, C. Hara, A., Hayashi, N. Hill, D., Fukunishi, Y., Furuno, M., Hanagaki, T., Hama, D., Imotani, K., Ishii, Y., Itoh, M., Iwama, K., Hiroaka, T., Hori, F., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Kasuura, T., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, S., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C. Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, D., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamahata, I., Yanai, S., Yanai, A., Yoshida, K., Yoshino, M., Yamatsuki, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN (GEL), RIKEN Genomic Science Center (GSC), RIKEN Yokohama Institute, 1-7-72 Shinhwa-cho, Tsukuba-shi, Ibaraki, 305-8565, Japan (E-mail: ygenome@rsc.riken.go.jp, URL: ygenome@rsc.riken.go.jp, Fax: 81-45-503-9216, Tel: 81-45-503-9222, Please visit: ygenome@rsc.riken.go.jp)

further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 15' GAGAGAGGAGAGATCCACAGACTCTTTTTTTTTTTTTTNN 3'. cDNA was
 prepared by using rnasolase thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot = 10.0 and subtraction to
 Rot = 60.0. Second strand cDNA was prepared with the primer adapter
 of sequence 15'
 GAGACGAGATTCGCGTTAATTAATTAATTCGCCCCCCCCCCC 3'. cDNA was cleaved
 with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
 bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3'
 end: BamHI. Host: DH10S.
 Location/Qualifiers
 1..3436

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MSD:MGI:1907060"
/clone="3230402E04"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="14, 17 days embryo"
1..3436
/gene="Sema3b"
1..3436
/misc_feature

```

```

/gene="Sema3b"
/note="data source:MGD, source key:MG1:107561,
evidence:ISS
sema domain, immunoglobulin domain (Ig), short basic
domain, secreted, (semaphorin) 3B"
BASE COUNT      754 a
ORIGIN           942 c 1001 g 739 t

```

Query Match	17.8%;	Score 466.8;	DB 11;	Length 3436;
Best Local Similarity	56.0%;	Pred. NO. 2.1e-83;		
Matches 1113;	Conservative	1;	Mismatches 768;	Total 1881

QY	356	CCGGCCGACAGTGTGCCCCGCTGGGCTCTCACTACCGAACCTCTGTCTGCCAACGCT	415
Db	226	CTGCCCTTAACCTTCCCGGCTTCGCTCTCTTCAAAATTACAG---GCCGGGATG	285
QY	416	CTGCCATCTTCTGGGCCCCCAGGGGTCCCTGACCTCCAGGCCATGTACCTTAATAGT	475
Db	283	GTGTCGAACCTTCAGGCTGACCGGACCTGCTGTATGAAAGCTTGCTGTGTATGAG	342
QY	476	ACCGAGACCCGCTCTTCTGAGGTGGCTGTGGACCGCCCTTACTCTCTGGGCTGACCAAG	535
Db	343	AGCGTGAAGCCCTGTTTGTGGGTGTGAACACAGTGTCTCCCTCAGCTGGACAA	402
QY	536	CATGCCAGATCCCCGGGAGTCTGTGGCCACCGACGACGAGGACGAGGAGAGTGTG	595
Db	403	TCAGCAAGGACCAAGAAAGCTGGCTGCGCCGCCCTGTGGATAGGCGTGAAGATGCA	462
QY	596	TTCGAAAGGAGAGATCTTGTGACAGATGCGCCAACTTCTGTCGGGTGTACAGCCTC	655
Db	463	ACTGGGACGAGGAGACATTTGTTACAGATGATGAATCTCTGTAAGTGTGTGACACT	522
QY	656	ACAACCGGACCCACTCTGTAGCCTGTGGACTGGGCTCTCCAGCCACCTGTGCTCA	715
Db	523	ACAACCAACCCACTCTGTGCTGGCTGTGGACAGAGGCTTTTCAACCCAACTGTGCTTTG	582
QY	716	TCACAGTTGGCCACCG---TGGGAGACATGTGTCCACTGGAGCTGGCAGTGTGAAA	772
Db	583	TGAGAGTGGGACACCGGCTGGAGAACCCATCTTCACTGTACCGGAGGAAACTTGAG	642
QY	773	GTGGCCCGGGGGCGGTGCGCTCAAGAACCCACGCTCCCTTGGACAGACCTTATACG	832
Db	643	ACGGCAAGGAGGAGACACTCTTATGACCCAAAGCAATCGGGCTGCTCGTGTGTGGGG	702
QY	833	GGGAGCGATGACAGGGTCTCACTGCTGACTTCTGGGGCGAGAGCCCATATCTTCCGA	892
Db	703	AAGACATGATTTCTGGGGTGAACAGACACTTATGGCGGGACTTTACATTTTTCGAA	762
QY	893	GTTGAGGCTCTTGGGACAGCTGTGCGTTCCGACCTGTACACAGTCTTGTG---ACGAC	949
Db	763	GCTGTGGTACAGATCCGAGTCTCGAACAAGCCCACTGATTCCTCGCTGCATATGAAC	822
QY	950	CCGGGTGTGTGGCCCGCCCGGATCCCTAGAACTGTACAGGACCAATGTGCAAGGT	1009
Db	823	CCAAATTTGTCAAGTCTTTTGGATCCCAAGAGATGAGAAACCTGTATGAGATTAATCT	882
QY	1010	ACTTCTCTTCTCGGAGAGCGTCCCTCGGCCCATGATGGCTCGAACATGTCACTGTCA	1069
Db	883	ATTCTCTTCTCGGAGTCCGCTGTGTAAGACACACAAATGGGGGCAATGTCTGTGT	942
QY	1070	GCGCGTGGGCGCGTCTGCTGAATGATGTGGGGGCGACAGGGGTGCTGTGTAACAAAT	1129
Db	943	CTCTGTTTGGCCAGATCTGAGGAAATGACCTTGGTGGACAGGAGCTTGTGTCAACAAAT	1002
QY	1130	GGAGCACTTTTCCAGAGGCGAGCTGTGCTCGTCCGTCGCCGCGCTGTGTGTGCGGAGA	1189
Db	1003	GGAGCAACTTTTCTGAAGAGGCGGCTTGTGTCTGATGACTGTGAGGT---GACA	1059
QY	1190	CCACATTTTACCAAGCTAAGATGTGTCTGCTGTGGGCCAAGCCGGGAGAGAGCCCTG	1249
Db	1060	CCACATTTTACCAACTTCAAGATGTTTCTTCTCTCCCGAGACCCCGACACCTTC	1119
QY	1250	AGGTGTACCGCTGTGTGACGACGCTCAAGTCCGCTGTTCAGAGGCTTGGCGCTGTGGT	1309
Db	1120	TTCCTGTATGTCTTCTTCCACCTCAAGTGTCTTTCAGAGGCTGTGTGTGTCTGTGT	1179
QY	1310	ACCCATTTTCCAGACATCTGGGAGGTTTTCAGCGGGCTTTGGCCACGGAATGGGCTTC	1369
Db	1180	ACAGATATGACGATGTGCGCGAGACCTTCTTGGAGACTTTTCTTACAAAGAGGCGCTTA	1239
QY	1370	AGCAACAGTGGGGCTTATGGGGGCAAGTCCCTTCCCTGCGCTGTGGCTGTGGCCCA	1429
Db	1240	CACACAGTGGGTGTCTTCCACAGGTCTGTCTCCCTACCAAGACCTGTGATGCGCCCA	1299
QY	1430	GCAAGATGACCGACACCAAGACGCGCTTTTGGCAGCAGCAAGGACTTACCGAGATGAG	1489

Db 127 GAAGCAGCAGCGAATCTTTCACACGAGGCGGGGCTGCTGTTCGGCAGGGTTAGCCG 186

QY 2220 TTTCGATGGGGGCACTACACCTGCACACACTGTGGAGCATGGCTTTCACAGACTGGGT 2279

Db 187 TTTCGATGGGGCACACTACACCTGCACACACTGTGGAGCATGGCTTTCACAGACTGGGT 246

QY 2280 CCGCGTGGCTCTGGTGGATGTTGGCTTCACAGACTGTTCACAGACTGGGT 2339

Db 247 CCGCGTGGCTCTGGTGGATGTTGGCTTCACAGACTGTTCACAGACTGGGT 2397

QY 2340 AAAGCCAGAGAGCCCGCCAGCCCGGGGAGGCGTGGCTTCACCCCAACCCAAAGGCTGG 364

Db 307 AAAGCCAGAGAGCCCGCCAGCCCGGGGAGGCGTGGCTTCACCCCAACCCAAAGGCTGG 364

ACCESSION	BI067451	GI:14474973
VERSION	BI067451.1	
KEYWORDS	EST	
SOURCE	chicken.	
ORGANISM	Gallus gallus	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Galliformes; Phasianidae;	
	Phasianidae; Galliformes; Phasianidae;	

Townsend Hall, Newark, DE 19713
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu
location@qualifiers

```

/db_xref="taxon:9031"
/clone="pf1n.pk012.d9"
/clone_lib="normalized chicken fat cDNA library"
/sex="Male and Female"
/tissue_type="fat"
/lab_host="E.coli EMDH10B"
/vector="pSPORI1"
/notes="104 t 26 others"

```

Query Match	Score	Score	Indels	Gaps
Best Local Similarity	12.2%	22.9e-54	1	1
Matches	72.6%	Pred. No. 2.9e-54	154	1108
Matches	411; Conservative	Mismatches	154;	

DY 1049 GCTCGNACCATGTACTGTCAGCCGGCAGGCC 62
 ||| | | | | | | | | | | | | | | | | |
 3 GCACGAGACACACCATCGTCAAGCCGGTGGCAGCGGCTCGGTAATGATGCTGGAGGCC 1168

1109 ACCGGGTCTGTTGAACAATGAGACACTTTCCTCAAGGCCAACCCCTCCTCTGTC
OY ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 63 AGAGGGTGTGGTCAACAGAGCGCGGCGC 1228
1169 CCGGCGCTGTGTGTGCCGAGACCCACTTGCACCACTAGAGGATGTTCTGCTGTGGC 182

123 CGGGTCTGGCGGCATTGATACCACTTTTGATGACCTGAGAGATGTCCTTTTACCACTTTT
 Db 1288

[illegible]

RESULT 6	436 bp	mrna	linear	EST 28-DEC-2001
BM287610				
LOCUS	BM287610	528378	MARC 3BOV Bos	taurus CDNA 5', mRNA sequence.
DEFINITION	BM287610			
ACCESSION	BM287610.1	GI:17996636		

VERSION	EST.
KEYWORDS	
SOURCE	COW
ORGANISM	
nos taurus	Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;	Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Mammalia; Eutheria;	
Bovidae; Bovinae; Bos;	
1 (bases 1 to 436)	
Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,	
White,J., Cho,J.J., Fahnstuck,S.C., Bennett	
A., Chittko-McKown,C.G.	

TITLE	JOURNAL	MEDLINE	COMMENT
Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle genome Res. 11 (4), 626-630 (2001)	21180013	Contact: Smith TTP, Animal Research Center, USDA, ARS, US Meat Research Center, NE 68933-0156, USA	

PO Box 166, Clay Center,
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.maric.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCTAGTCACGAC
Plate: 137 row: N column: 23
Seq primer: APTTAGTGACACTATAG.
Location/Qualifiers

	FEATURES	SOURCE
62	1. .436 locus="9913" /organism="Bos taurus" /db_xref="txon:9913" /clone_id="MARC 3BOV" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI" library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and f	a 161 c 131 g 82 t

Thu Oct 10 09:35:02 2002

us-09-813-290-1.1st

FEATURES
Source Location/Qualifiers
1. 785
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:535609"
/clone_1bp="NCI.CGAP_Mame"
/sex="female, virgin"
/false_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/dev_host="MD10B"
/lab_host="MD10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site: 1; Salt:
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 161 a 239 c 185 t
ORIGIN 10.1%; Score 264.2; DB 10; Length 785;
Query Match 62.3%; Pred. No. 5.8e-43; Indels 17; Gaps 4;
Best Local Similarity 0; Mismatches 278;
Matches 488; Conservative
1064 CTGTGACCCGCTGGCGCGCTGCTGTA-TGATCTGGGGGCGGCGCTGCTG 1122
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
9 CTGTGCTGCTGTGGCGAGATCTCAGAGAGTACCTGGTGGCGGCGCTGGTGT 1182
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1123 AACAAATGACACTTCTCTCAAGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTG 1242
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
69 AACAAATGACACTTCTCTCAAGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTG 185
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1183 GCCGAGACACTTCTCTCAAGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1302
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
129 ---GACACCACTTCTCTCAAGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1243 AGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1362
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
186 ACACCTCTTCT 305
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1303 TGTGTATACATGTCACATGTCAGGAGGCTTCTCAAGGCGAGGCTTCTCTCTCTCTCT 1422
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
246 TGGCGTACACATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1422
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1363 GGGCGTACACATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 365
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
306 GGGCGTACACATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1482
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1423 TGGCGTACACATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 413
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
366 TGGCGTACACATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1542
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1483 GATGAGTGTCTGATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 473
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
414 GATGAGTGTCTGATGTCAGGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1602
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1543 CATGCGCGCGCTGCT 533
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
474 GGGGGGCGCGCT 1652
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1603 GAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 593
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
534 GAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4722
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1663 TGTGTCT 653
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
594 ACAGTGTCT 1782
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1723 CTGAGAGGCT 713
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
654 CTGAGAGGCT 1841
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
1783 AAAGGCAATCT 773
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
714 AAAGGCAATCT

1842 ATG 1844
11
Db 774 CTG 776
ORIGIN 91 a 210 c 193 g 112 t 3 others
BASE COUNT 91 a 210 c 193 g 112 t 3 others
Query Match 9.7%; Score 255.2; DB 10; Length 609;
Best Local Similarity 67.9%; Pred. No. 3.4e-41; Indels 6; Gaps 2;
Matches 385; Conservative
354 CCG 473
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
43 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 162
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
414 CTCTGCGATCT 533
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
103 CTCTGCGATCT 222
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
474 GTACGAGAGCGCGCT 593
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
163 GTACGAGAGCGCGCT 282
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
534 GCGATGCGAGAGCGCGCT 653
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
223 GCGATGCGAGAGCGCGCT 342
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
594 TGTGTGAG 713
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
283 TTTTTCGAG 402
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
654 TCACACGAG 402
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
343 CTACACGAG

FEATURES
Source
1. 609
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pGf1n.pK007.m15"
/clone_1bp="normalised chicken fat cDNA library"
/sex="Male and Female"
/false_type="fat"
/lab_host="E. coli EMDH10B"
/note="Vector: pSPORT1"
Location/Qualifiers
1. 609 bp mRNA linear EST 15-JUN-2001
BIO66327
LOCUS
DEFINITION
pGf1n.pK007.m15 normalized chicken fat cDNA library
cdna clone pGf1n.pK007.m15 5' similar to g11910362
ref|NP_064548.1| semaphorin sem2 [Homo sapiens] db|BA98132.1|
ref|XP_003233.1| semaphorin sem2 [Homo sapiens] g, mRNA sequence.
(AB029496) semaphorin sem2 [Homo sapiens]
BIO66327
BIO66327.1 GI:14473849
EST.
Chicken.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Ardeosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 609)
Cognburn, L.A., Morgan, R.W. and Burnside, J.
Chicken ESTs from fat
Unpublished (2001)
Contact: Larry A. Cognburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cognburn@udel.edu, www.chickest.udel.edu.
Location/Qualifiers

OY 714 CATCAGAGTTGGCCACCGTGGAGCATGTG-----CTCCACCTGAGAGCTTGGCAGTGTG 768
 Db 403 CATCTACGTGGGGGACCGGGGAGACCCAGCATGTGCTTACGCTTGAGACCTGGTACAG 462
 OY 769 GAAATGAGCGGGGGCGGTGCTGCTCAGACGCCAGCCGCTCCCTTGGCCAGCACTTCACT 828
 Db 463 GAGAGTGTGCTGCGAGAGTGGCCACATGACCGCTGACATTTTGGCCAGCANNATCAT 522
 OY 828 AGAGCGAGAGCTGTACACGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 887
 Db 523 CGGTGGGAGAGCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 887
 OY 888 CCGAAGTGGAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914
 Db 583 CCGAGCAGCAGCGGAGCCCGCGCGCT 609

RESULT 10
 LOCUS B1652258
 DEFINITION 603299702F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5340217 5'
 mRNA sequence.
 VERSION B1652258
 KEYWORDS EST
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 990)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@email.nih.gov
 Tissue Procurement: Lohar Hemmighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNL at:
<http://image.lnl.gov>
 Plate: LLM1864 Row: C Column: 02
 High quality sequence stop: 892.

FEATURES
 source Location/Qualifiers
 1. 990

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="5340217"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORE; Site: 1; NotI; Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lohar Hemmighausen/Chu-Xia Deng, NIH Reference (1999). Note: this is a NCI-CGAP Library."
 194 a 285 c 292 g 219 t

Query Match 9.5%; Score 250.2; DB 10; Length 990;
 Matches 514; Conservative 0; Mismatches 338; Indels 11; Gaps 5;

OY 577 GGACAGAGGAGAGTGTTCGAAGAGAGATCTTTGACAGAGTGGCGCACTTC 636
 Db 71 GAATGGCTGTGAAGATGCACTGGGAGAGAGATTTGGTACGAGTGAATCTTC 130
 OY 637 GTGGCGGCTGACAGCTTACACCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 696
 Db 131 GTGAAGCTGTGACAGCTTACACCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 190

OY 697 CAGCCACCTGTGCTGCTATCATCAAGTGGCCAGCC---TGGGAGCATGTGCTCACCTG 753
 Db 191 CACCAACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
 OY 754 GAGCTGCGCATGTGTGAAGAGTGGCGGGGGGCTGCTGCTGCTGCTGCTGCTGCT 250
 Db 251 GAGCTGCGCATGTGTGAAGAGTGGCGGGGGGCTGCTGCTGCTGCTGCTGCTGCT 813
 OY 814 GCGACACCTTATGACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310
 Db 311 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 873
 OY 874 GAGCCATGATGCTTCCAGAGTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370
 Db 371 GACTTATCATGCTTTCGAGAGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 933
 OY 934 AGTCTGCTGCT---AGACCCCGGCTTGTGATGGCGCCCGATCCCTGAGACCTGAT 430
 Db 431 TCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 990
 OY 991 CAGGACATGACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
 Db 491 CCTGATGACATTAATAATCTATTTCTTCCCGAGTCCGCTGCTGCTGCTGCTGCT 1050
 OY 1051 TCGAACCATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
 Db 551 ATGGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
 OY 1111 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610
 Db 611 CGGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
 OY 1171 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
 Db 671 GAGTGTGAGCT---GACACCCACTTGTGACCACTGACAGTGTGCTGCTGCTGCT 1230
 OY 1231 AAGCCCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
 Db 728 CGAGACCGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1290
 OY 1291 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
 Db 788 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1349
 OY 1350 TCCCGACGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
 Db 848 TGGCTTACAGAGAGGCTTACACCGAGTGGTGTCTTACCCCGGGTGGGTTCTTATTC 1409
 OY 1410 TCGCCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
 Db 908 CAGACCTGG-ATGTGCCAGCA 929

RESULT 11
 LOCUS B1557438
 DEFINITION 603238962F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5292039 5'
 mRNA sequence.
 VERSION B1557438
 KEYWORDS EST
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 662)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@email.nih.gov
 Tissue Procurement: Lohar Hemmighausen Ph.D., Chu-Xia Deng Ph.D.

Query Match 9.5%; Score 250.2; DB 10; Length 990;
 Matches 514; Conservative 0; Mismatches 338; Indels 11; Gaps 5;


```

OY 711 CTTATCATGAGTTGGCCACCG---TGGGAGCATGTGCTCCACCTGAGGCTGGAGCTGT 767
Db 457 CTTTGTGAGAGTGGCCACCGGGGAGAGAGCCGCTCTCCGCTGACCCAGAGAGAT 516
OY 768 GGAAGTGGCCGGGGGGGGTGGCTTCAGCAGAGCCCGCTCCCTTTCAGCAGCTTCAT 827
Db 517 AGAGATGTGAGGAGGAGGAGAGAGTCTATGACCCAGAGATGGGGTCCCTCCGCTGT 576
OY 828 AGACGGGAGAGTGTACAGAGGCTCTACCTGTCTCTGTGGGGGAGGAGGAGCTATGCT 887
Db 577 GGGGAGAGAGTATCTACTAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636
OY 888 CCGAGTGGAGAGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 636
Db 637 TCGCAGGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
OY 945 CGACCCCGGCTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Db 697 TGAGCCCAAGTGTCTCAAGATATTTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
OY 1005 GGTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1064
Db 757 AATCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 816
OY 1065 TGTCAAGCCGCTGGGCGCCGCTGTGCTGAGATGATGCTGGGGGAGAGAGAGAGAGAG 816
Db 817 CGTGTCCCGCTGGGCGCCGCTGTGCTGAGATGATGCTGGGGGAGAGAGAGAGAGAGAG 1124
OY 1125 CAATGAGACACTTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
Db 876 CAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 920

```

```

RESULT 13
AL584324
LOCUS 786 bp mRNA linear EST 28-FEB-2001
DEFINITION Gallus gallus cDNA clone ROS007F06, mRNA sequence.
ACCESSION AL584324.1 GI:13163055
VERSION 1
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
1 (bases 1 to 786)
Murray, F.
Stratagene Chick Embryo Lambda cDNA Library
Unpublished (2001)
Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
Seq primer: T3.

```

```

FEATURES
source
1..786
Location/Qualifiers

```

```

/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="ROS007F06"
/clone_11b="Stratagene Chick Embryo Lambda cDNA Library (*
937405)"
/tissue_type="Embryo"
/dev_stage="5 days old"
/lab_host="SOLR cells (kanamycin resistant)"
/Note="Vector: pBLUESCRIPT SK; Site 1: EcoRI, Site 2: XhoI
vector. Average insert size: 1.5kb.; 5' adaptor sequence:
5' GAATTCGCGACGAG 3'; 3' adaptor sequence: 5'

```

```

BASE COUNT 166 a 242 c 218 g 155 t 5 others
ORIGIN

```

```

Query Match 9.0%; Score 237.4; DB 9; Length 786;
Best Local Similarity 61.7%; Pred. No. 1,4e-37;
Matches 434; Conservative 0; Mismatches 253; Indels 16; Gaps 3;

```

```

OY 960 GATGCGCCCGGAGTCCCTGAGAACTCTGACCAAGAGAGAGAGAGAGAGAGAGAGAG 1019
Db 1 GAATTCGCGACGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
OY 1020 CTGGAGACAGCTCCCTGCGCCGATGAGTGGTGAACCAATGCTGTCAGAGAGAGAGAG 1079
Db 61 CCGGAGAGAG-----TCGGGAGATGCCCGGCTCACCGCCGGGGGTCTATTCGCGATGG 114
OY 1080 CCGGCTGCGTGAATGATGCTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
Db 115 GCGCATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 174
OY 1140 CTTCAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1199
Db 175 CTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1234
OY 1200 CCAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1259
Db 235 TGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
OY 1260 GCTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1319
Db 295 GCTTCTCTCCGATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 354
OY 1320 AGACATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1379
Db 355 CGACATCCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
OY 1380 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1439
Db 415 GATGCGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474
OY 1440 CGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1499
Db 475 CACACCTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525
OY 1500 TCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1559
Db 526 CATGCGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
OY 1560 TGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619
Db 586 CTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644
OY 1620 GGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662
Db 645 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

```

RESULT 14
AL584625
LOCUS 754 bp mRNA linear EST 28-FEB-2001
DEFINITION Gallus gallus cDNA clone ROS012B12, mRNA sequence.
ACCESSION AL584625.1 GI:13163356
VERSION 1
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
1 (bases 1 to 754)
Murray, F.
Stratagene Chick Embryo Lambda cDNA Library

```


Db 241 GACTTACCATTTGCAAGCCCTTGATGAGATCTCCGACAGAGCCCAATGA 300
 QY 933 GAGTCTCTTGC---ACGACCCCGGTTGTGATGAGCCCGGATCCCTGAGAACTCTGA 989
 Db 301 TTCCCGCTGCTCAATGAAACCAAGTTGTCAAGSTCTTTGATCCCAAGAGGTGAGAA 360
 QY 990 CAGAGCAATGACAAAGGTGATCTTCTTCTCGAGACGCTCCCTCGCCGATGATG 1049
 Db 361 CCTGATGACGATAAATCTATTCTTCTTCTCGGAGTCCGCTGTGGAAGCAGACACAGC 420
 QY 1050 CTCGACCATGTCTACGTGTCAGCCGCTGAGCCGCTGTGATGATGATGAGGAGCA 1109
 Db 421 AATGGGGCGGATGCTGTCTGTCTGTGTCAGATCTGAGAAATGACCTGGGTGGCA 480
 QY 1110 GCGGTGCTGTGACAAATGAGCACTTCTCAGAGCCAGCTGTGCTGCTGCTGCC 1169
 Db 481 GCGGAGCTTGTGTCACAAATGAGCACTTCTGAAAGCGCGCTGTGTGCTCAGTACC 540
 QY 1170 CGGCTGTGTGTCGAGAGCCACTTGTGACCAAGTATGATGCTGTGCTGAGCC 1229
 Db 541 TGGAGTTGAGGT---GACACCCACTTGTGACCACTTCAGATGTTTCTTCTGCTC 597
 QY 1230 CAAGCGCGGAGAGCTGAGAGTGTACGCGCTGTGACAGCCGTGATGAGCTGTTCA 1289
 Db 598 CCGAGACCGCCAGACCTCTTCTGTATGCTGTCTTCCACTCCAGTGTCTTCCA 657
 QY 1290 GGGCTTCCGCTGTGTGTATGACATGAGGAGATCTGGAGAGTTTCAACGGCCCTT 1349
 Db 658 GGGCTTGT 716
 QY 1350 TGGCCACGAGATGAGGCTCTAGCAGCAGTGGGGCCCTATGAGGAGCAAGTGGCCCTTCC 1409
 Db 717 TGTCTACAAAGAGGAGGCTCTACACAGAGTGGGTGTCTTACAGAGGTGTGTCTTAAAC 776
 QY 1410 TCGCCCTGGGCTGTGAGCCCAAGATGACCGCAGACAGGAGAC 1453
 Db 777 AAG-ACCTGAGATGTGCCCAAGAGACCTTGGCACCTTCAAGTTC 819

Search completed: October 9, 2002, 13:14:51
 Job time : 2034 secs

Thu Oct 10 09:35:02 2002

us-09-813-290-2.rag

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 12:40:53 : Search time 59 Seconds
(without alignments)
1647.281 Million cell updates/sec

Title: US-09-813-290-2

Sequence: 1 MACALAGKVFPMGSPVWVK.....KMSRVHAENHRTPREVAT 875

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

747574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

A: Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the total score distribution.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4746	100.0	875	22	AA65619 Novel human protei
2	4215	88.8	782	21	AA65619 Novel human protei
3	4215	88.8	782	22	AA65619 Novel human protei
4	4203.5	88.6	785	21	AA65619 Novel human protei
5	4046.5	85.3	779	22	AA65619 Novel human protei
6	2052.5	43.2	775	19	AA65619 Novel human protei
7	2041.5	43.0	775	20	AA65619 Novel human protei
8	1942.5	40.9	777	20	AA65619 Novel human protei
9	1942.5	40.9	777	22	AA65619 Novel human protei
10	1830	38.6	771	16	AA65619 Novel human protei
11	1830	38.6	771	22	AA65619 Novel human protei

12	1810	38.1	796	19	AA65619 Novel human protei
13	1747	36.8	749	22	AA65619 Novel human protei
14	1721	36.3	777	20	AA65619 Novel human protei
15	1721	36.3	777	21	AA65619 Novel human protei
16	1721	36.3	777	22	AA65619 Novel human protei
17	1721	36.3	777	22	AA65619 Novel human protei
18	1686	35.5	807	22	AA65619 Novel human protei
19	1652	34.8	751	22	AA65619 Novel human protei
20	1540	32.4	751	20	AA65619 Novel human protei
21	1538	32.4	751	21	AA65619 Novel human protei
22	1538	32.4	751	22	AA65619 Novel human protei
23	1289	27.2	861	18	AA65619 Novel human protei
24	879.5	18.5	833	22	AA65619 Novel human protei
25	876	18.5	861	19	AA65619 Novel human protei
26	874.5	18.4	861	22	AA65619 Novel human protei
27	874.5	18.4	861	22	AA65619 Novel human protei
28	867.5	18.3	862	18	AA65619 Novel human protei
29	867.5	18.3	862	22	AA65619 Novel human protei
30	867.5	18.3	862	22	AA65619 Novel human protei
31	867.5	18.3	862	22	AA65619 Novel human protei
32	862	18.2	834	22	AA65619 Novel human protei
33	862	18.2	834	22	AA65619 Novel human protei
34	855	18.0	837	21	AA65619 Novel human protei
35	855	18.0	837	22	AA65619 Novel human protei
36	855	18.0	837	22	AA65619 Novel human protei
37	851.5	17.9	590	22	AA65619 Novel human protei
38	851.5	17.9	590	22	AA65619 Novel human protei
39	851.5	17.9	590	22	AA65619 Novel human protei
40	824.5	17.4	770	22	AA65619 Novel human protei
41	776	16.4	770	21	AA65619 Novel human protei
42	751	15.8	776	19	AA65619 Novel human protei
43	750.5	15.5	761	21	AA65619 Novel human protei
44	737	15.5	761	21	AA65619 Novel human protei
45	737	15.5	761	21	AA65619 Novel human protei

ALIGNMENTS

RESULT 1	AA65619	standard; protein; 875 AA.
AC	AA65619	
DT	07-JAN-2002	(first entry)
DE	Novel human protein (NHP) sequence.	
XX	NHP: novel human protein; secreted protein; semaphorin; oxytocin;	
XX	neurohypophysis; neuropeptide; gene therapy; drug screening.	
OS	Homo sapiens.	
XX	WO200170806-A2.	
XX	27-SEP-2001.	
XX	20-MAR-2001: 2001WO-US08834.	
XX	20-MAR-2000: 2000US-190638P.	
XX	22-MAR-2000: 2000US-191188P.	
XX	31-MAR-2000: 2000US-193639P.	
XX	(LEXI-) LEXICON GENETICS INC.	
XX	Turner CA, Hilbun E, Wang X, Donoho G;	
XX	Walke DW, Wilgowski NL,	
XX	Scoville J;	
XX	WPI: 2001-611483/70.	
XX	N-PsDB: AA47790.	
XX	New polynucleotides encoding human proteins that share structural	
PT		

PT similarity with semaphorin proteins, protein hormones of
PT neurohypophyseal family for drug screening, diagnosis and therapy of
XX biological disorders
PS Claim 2: Page 35-37; 43pp; English.

CC The invention relates to novel human secreted proteins (NHP) that share
CC structural similarity with semaphorin proteins, protein hormones
CC of the neurohypophyseal family and oxytocin (neurophysin I precursor)
CC techniques for treating symptomatic or phenotypic manifestations of
CC encoding NHP products are useful in drug screening
CC expression. The constructs are useful in gene therapy for modulating NHP
CC function as bioreactors in vivo, these genetically engineered host cells
CC also find use in molecular mutagenesis/evolution of proteins that are
CC partially encoded by the NHP sequences. The encoded NHP polypeptides are
CC useful for generating antibodies, as reagents in diagnostic assays, for
CC identifying other cellular gene products related to NHP and as reagents
CC in assays for screening for compounds that are useful in the treatment of
CC mental, biological or medical disorders and diseases. The present
CC sequence represents the amino acid sequence of a NHP.

Sequence 875 AA:

Query Match 100.0%; Score 4746; DB 22; Length 875;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 875; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MACALAGKVFPGSPVWVKSLHMANKEGAGRGSGFLLSSAPLPAQDWEPLPK 60
QY 61 WMPGSSRANVRNRRPAGGEGSAGRRORCPFPSPASAMACWLLGGLHGGSSGSPG 120
DB 61 WMPGSSRANVRNRRPAGGEGSAGRRORCPFPSPASAMACWLLGGLHGGSSGSPG 120
QY 121 PSVRLRLSYRDLISANRSALFPGPSLNTLQAMLYDEYRDLFGGLDLYSLRLDQAM 180
DB 121 PSVRLRLSYRDLISANRSALFPGPSLNTLQAMLYDEYRDLFGGLDLYSLRLDQAM 180
QY 181 PDREVYLPMPGQRECEYKRGDRPLTECANFVRLDPRHNTHLACGTGAFPTCALIT 240
DB 181 PDREVYLPMPGQRECEYKRGDRPLTECANFVRLDPRHNTHLACGTGAFPTCALIT 240
QY 241 VGRHEVHLHLEPGSVESGRCRCPHPSRPFSTIDGLYTLGTLADFLGRLMFLRSGG 300
DB 241 VGRHEVHLHLEPGSVESGRCRCPHPSRPFSTIDGLYTLGTLADFLGRLMFLRSGG 300
QY 301 PRPALRSDSDSLHDPFRFMAARIPENSDDNDKYFFESFETVSPGGSHVTVSVG 360
DB 301 PRPALRSDSDSLHDPFRFMAARIPENSDDNDKYFFESFETVSPGGSHVTVSVG 360
QY 361 RVCVNDAGGQRLVYNKSTFLARLVCSVPGFAGETFDLEVFLLMPKAGSLEVA 420
DB 361 RVCVNDAGGQRLVYNKSTFLARLVCSVPGFAGETFDLEVFLLMPKAGSLEVA 420
QY 421 LFTSVAVPQGFVYHMAIDWEVNGFPAHRDGRPHONGPFGKVPFPCVSKMT 480
DB 421 LFTSVAVPQGFVYHMAIDWEVNGFPAHRDGRPHONGPFGKVPFPCVSKMT 480
QY 481 AOPGPRGSTKDYDEVLQFARAHPLMFWRPRRGRPLVTKHQAOLHQLVTVRVAE 540
DB 481 AOPGPRGSTKDYDEVLQFARAHPLMFWRPRRGRPLVTKHQAOLHQLVTVRVAE 540
QY 541 DGYTDVFLGTGDSGVLYKALQAGSAPDEEVLELQVFPVPTTEMEISVKKRLY 600
DB 541 DGYTDVFLGTGDSGVLYKALQAGSAPDEEVLELQVFPVPTTEMEISVKKRLY 600
QY 601 VGSRLGVAOLRLHOCCTYGTACACCLARDPYCAMDASCHYRPSLGRFRRODRIHG 660
DB 601 VGSRLGVAOLRLHOCCTYGTACACCLARDPYCAMDASCHYRPSLGRFRRODRIHG 660

DB 601 VGSRLGVAOLRLHOCCTYGTACACCLARDPYCAMDASCHYRPSLGRFRRODRIHG 660
QY 661 NPALQCLGSGEAEVAGVLAATWYGTENSTFLECKPSQAAVRLQRPEDGPDQV 720
DB 661 NPALQCLGSGEAEVAGVLAATWYGTENSTFLECKPSQAAVRLQRPEDGPDQV 720
QY 721 KTDERVLTERGLFRRISRPDACTYCTLEHFGSQVRLALVYASQDLNLPPEP 780
DB 721 KTDERVLTERGLFRRISRPDACTYCTLEHFGSQVRLALVYASQDLNLPPEP 780
QY 781 KPEPPRAGGLASTPPRMYKDIOLIGFANLPRDVEYERWCGTSGCFRRSRG 840
DB 781 KPEPPRAGGLASTPPRMYKDIOLIGFANLPRDVEYERWCGTSGCFRRSRG 840
QY 841 KQARKSMAGLELGGKMKSRVHAENHRTPREVEAT 875
DB 841 KQARKSMAGLELGGKMKSRVHAENHRTPREVEAT 875

RESULT 2

AA23609
ID AAB23609 standard; Protein: 782 AA.

AC AAB23609;
DT 12-JAN-2001 (first entry)

DE Human secreted protein SEQ ID NO: 18.
XX

KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
XX haematopoiesis regulation; tissue growth; haemostasis; inflammation.

OS Homo sapiens.
XX

Key
FH Peptide
FT 10..22
FT /label= signal_peptide
FT 23..782
XX /label= mature_protein

PN MO2000049134-A1.
XX

PD 24-AUG-2000.
XX

PF 18-FEB-2000; 2000MO-US04340.
XX

PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 15-FEB-1999; 99US-0167823.
XX 15-FEB-2000; 2000US-0298733.

PA (ALPH-) ALPHAGENE INC.
XX

PI Valenzuela D, Yuan O, Hoffman H, Hall J, Raplejo P.
XX
DR WPI: 2000-549267/50.
XX
DR N-PSDB: AAA3109.

PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
XX as well as nutritional sources or supplements
PS Claim 27: Page 250-253; 309pp; English.

CC The present sequence is the sequence of a human secreted protein. Its
CC cDNA was isolated from an adult brain cDNA library. The proteins
CC and coding sequences of the invention can be used in the isolation of
CC similar genes and proteins, in the elucidation of their function in vivo,

Thu Oct 10 09:35:02 2002

us-09-813-290-2.1rag

CC and to treat a number of conditions. It is possible that they may have
 CC uses as nutritional supplements, as cytokine or cell proliferation
 CC factors, in immune modulation, where they may be used to treat immune and
 CC autoimmune diseases, as haematopoietic regulators (treating myeloid or
 CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
 CC have chemokine or chemotactic activity, haemostatic or thrombolytic
 CC activity, or anti-inflammatory activity.

XX Sequence 782 AA:

Query Match 88.8%; Score 4215; DB 21; Length 782; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Mismatches 0;

Matches 782; Conservative 0;

94 MAPSAMATICMLLGGILLHGGSSGSPSPVRLSLSTRDLSSANSAIFLPGQSLNLOA 153
 1 MAPSAMATICMLLGGILLHGGSSGSPSPVRLSLSTRDLSSANSAIFLPGQSLNLOA 60
 154 MYLDEYRDRLFLGGLDALYSRLDQAMPDPREVLPMPQPGRECVKRGKRDPLTECANFV 120
 61 MYLDEYRDRLFLGGLDALYSRLDQAMPDPREVLPMPQPGRECVKRGKRDPLTECANFV 273
 214 RYLOPHNRTHLLACGTAFOPTCALITVGRGHEVHLLEPGSVESGRCGRCPHEPSRPFAS 180
 121 RYLOPHNRTHLLACGTAFOPTCALITVGRGHEVHLLEPGSVESGRCGRCPHEPSRPFAS 333
 274 TFLDGLYGLTADFLGREAMIFRSGGPRPALRSDSDSLHDPREFYMAARIPENSDDN 240
 181 TFLDGLYGLTADFLGREAMIFRSGGPRPALRSDSDSLHDPREFYMAARIPENSDDN 393
 334 DKYFFEFSETPSPDGSNNHVTYSRGRVCVNDAGGQRYLVNKNSTFLKALVCSVPQPG 400
 241 DKYFFEFSETPSPDGSNNHVTYSRGRVCVNDAGGQRYLVNKNSTFLKALVCSVPQPG 453
 394 GAETHFDLDEVLFLMKRAGKSLFYALFSTVSAYVQGFAYCYHMADINEVFNNGEPFHR 360
 301 GAETHFDLDEVLFLMKRAGKSLFYALFSTVSAYVQGFAYCYHMADINEVFNNGEPFHR 513
 454 DGPORHMPGGYGGKVPFRPGVCPSPKMTAQGPFRPGSTKDPDEVLOFARAHPLMFVPR 420
 361 DGPORHMPGGYGGKVPFRPGVCPSPKMTAQGPFRPGSTKDPDEVLOFARAHPLMFVPR 573
 514 RHGRPVLVKTHLAQOLHOIVYRVAEADGTVDYIFLGTDSGVYKVIALQAGSABPEEV 480
 421 RHGRPVLVKTHLAQOLHOIVYRVAEADGTVDYIFLGTDSGVYKVIALQAGSABPEEV 633
 574 VLEELQYFKVPTPTTEMEISVKRQMLYVGSRLGVAQLRHOCETGYTACAECLARDPYC 540
 481 VLEELQYFKVPTPTTEMEISVKRQMLYVGSRLGVAQLRHOCETGYTACAECLARDPYC 693
 634 AMDGASCHYRPSLGKRRFRRODIRGNPALQCLQSQSEAEVGLVAATMVYGEHNSTF 600
 541 AMDGASCHYRPSLGKRRFRRODIRGNPALQCLQSQSEAEVGLVAATMVYGEHNSTF 753
 694 LECLEKSPQAAVRLTORPDEGPDQVKTDERVLATERGLLFRLSRFDGTYCTTLEH 660
 601 LECLEKSPQAAVRLTORPDEGPDQVKTDERVLATERGLLFRLSRFDGTYCTTLEH 813
 754 GFSQTVRLALVIVASQDLNLFPEPEKPEEPAPAGLSTPPKAWYDILQIGFANLP 720
 661 GFSQTVRLALVIVASQDLNLFPEPEKPEEPAPAGLSTPPKAWYDILQIGFANLP 873
 814 RVDEYCEKRWCRGTTEGSCGFRSRSGKQARQKSNAGLELGKMSRYVAHEHNTPREVE 780
 721 RVDEYCEKRWCRGTTEGSCGFRSRSGKQARQKSNAGLELGKMSRYVAHEHNTPREVE 873
 874 AT 875
 781 AT 782

RESULT 3
 AAG65620

ID AAG65620 standard; Protein: 782 AA.

XX AAG65620;

XX 07-JAN-2002 (first entry)

XX Novel human protein (NHP) sequence.

XX NHP: novel human protein; secreted protein; semaphorin; oxytocin;

XX neurohypophyseal; nontropic; gene therapy; drug screening.

XX Homo sapiens.

XX WO200170806-A2.

XX 20-MAR-2001; 2001WO-050834.

XX 20-MAR-2000; 2000US-190638P.

XX 22-MAR-2000; 2000US-191188P.

XX 31-MAR-2000; 2000US-193639P.

XX (LEXI-) LEXICON GENETICS INC.

XX Walke DW, Wilganowski NL, Turner CA, Hilbun E, Wang X, Donoho G;

XX Scoville J;

XX WPI: 2001-611483/70.

XX N-PSDB: AAK47791.

XX New polynucleotides encoding human proteins that share structural

XX similarity with semaphorin proteins, protein hormones of

XX neurohypophyseal family for drug screening, diagnosis and therapy of

XX biological disorders

XX Claim 4; page 38-40; 43pp; English.

XX The invention relates to novel human secreted proteins (NHP) that share

XX structural similarity with semaphorin proteins, protein hormones of

XX the neurohypophyseal family and oxytocin (neurophysin 1 precursor)

XX family. The NHP nucleotide sequences are useful in drug screening

XX techniques for treating symptomatic or phenotypic manifestations of

XX NHP in the body. Nucleotide constructs

XX encoding the normal function of NHP in the body. Nucleotide constructs

XX encoding NHP products are useful in gene therapy for modulating cells

XX expression. The constructs can be used to genetically engineer host cells

XX to express NHP products in the body delivering a continuous supply of a

XX function as bioreactors in the body delivering a continuous supply of a

XX NHP, a NHP peptide, or a NHP fusion protein to the body. The sequences

XX also find use in molecular mutagenesis/evolution of proteins that are

XX partially encoded by the NHP sequences. The encoded NHP polypeptides are

XX useful for generating antibodies, as reagents in diagnostic assays, for

XX identifying other cellular gene products related to NHP and as reagents

in assays for screening for compounds that are useful in the treatment of

mental, biological or medical disorders and diseases. The present

sequence represents the amino acid sequence of a NHP.

XX Sequence 782 AA:

Query Match 88.8%; Score 4215; DB 22; Length 782; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Mismatches 0;

Matches 782; Conservative 0;

94 MAPSAMATICMLLGGILLHGGSSGSPSPVRLSLSTRDLSSANSAIFLPGQSLNLOA 153

1 MAPSAMATICMLLGGILLHGGSSGSPSPVRLSLSTRDLSSANSAIFLPGQSLNLOA 60

154 MYLDEYRDRLFLGGLDALYSRLDQAMPDPREVLPMPQPGRECVKRGKRDPLTECANFV 120

61 MYLDEYRDRLFLGGLDALYSRLDQAMPDPREVLPMPQPGRECVKRGKRDPLTECANFV 273

214 RYLOPHNRTHLLACGTAFOPTCALITVGRGHEVHLLEPGSVESGRCGRCPHEPSRPFAS 180

121 RYLOPHNRTHLLACGTAFOPTCALITVGRGHEVHLLEPGSVESGRCGRCPHEPSRPFAS 333

274 TFLDGLYGLTADFLGREAMIFRSGGPRPALRSDSDSLHDPREFYMAARIPENSDDN 240

181 TFLDGLYGLTADFLGREAMIFRSGGPRPALRSDSDSLHDPREFYMAARIPENSDDN 393

334 DKYFFEFSETPSPDGSNNHVTYSRGRVCVNDAGGQRYLVNKNSTFLKALVCSVPQPG 400

241 DKYFFEFSETPSPDGSNNHVTYSRGRVCVNDAGGQRYLVNKNSTFLKALVCSVPQPG 453

394 GAETHFDLDEVLFLMKRAGKSLFYALFSTVSAYVQGFAYCYHMADINEVFNNGEPFHR 360

301 GAETHFDLDEVLFLMKRAGKSLFYALFSTVSAYVQGFAYCYHMADINEVFNNGEPFHR 513

454 DGPORHMPGGYGGKVPFRPGVCPSPKMTAQGPFRPGSTKDPDEVLOFARAHPLMFVPR 420

361 DGPORHMPGGYGGKVPFRPGVCPSPKMTAQGPFRPGSTKDPDEVLOFARAHPLMFVPR 573

514 RHGRPVLVKTHLAQOLHOIVYRVAEADGTVDYIFLGTDSGVYKVIALQAGSABPEEV 480

421 RHGRPVLVKTHLAQOLHOIVYRVAEADGTVDYIFLGTDSGVYKVIALQAGSABPEEV 633

574 VLEELQYFKVPTPTTEMEISVKRQMLYVGSRLGVAQLRHOCETGYTACAECLARDPYC 540

481 VLEELQYFKVPTPTTEMEISVKRQMLYVGSRLGVAQLRHOCETGYTACAECLARDPYC 693

634 AMDGASCHYRPSLGKRRFRRODIRGNPALQCLQSQSEAEVGLVAATMVYGEHNSTF 600

541 AMDGASCHYRPSLGKRRFRRODIRGNPALQCLQSQSEAEVGLVAATMVYGEHNSTF 753

694 LECLEKSPQAAVRLTORPDEGPDQVKTDERVLATERGLLFRLSRFDGTYCTTLEH 660

601 LECLEKSPQAAVRLTORPDEGPDQVKTDERVLATERGLLFRLSRFDGTYCTTLEH 813

754 GFSQTVRLALVIVASQDLNLFPEPEKPEEPAPAGLSTPPKAWYDILQIGFANLP 720

661 GFSQTVRLALVIVASQDLNLFPEPEKPEEPAPAGLSTPPKAWYDILQIGFANLP 873

814 RVDEYCEKRWCRGTTEGSCGFRSRSGKQARQKSNAGLELGKMSRYVAHEHNTPREVE 780

721 RVDEYCEKRWCRGTTEGSCGFRSRSGKQARQKSNAGLELGKMSRYVAHEHNTPREVE 873

874 AT 875

781 AT 782

```

Db 121 RVLPNHNTHLACGTGAFOTCALITVGHREHVLHLEPGSVESGRGRCHEPSPFAS 180
QY 274 TFIIDGLYGLTADFLGREAMIFRSGGPRPALRSDDSLHDPFVMAARIPENSODN 333
Db 181 TFIIDGLYGLTADFLGREAMIFRSGGPRPALRSDDSLHDPFVMAARIPENSODN 333
QY 334 DKVYFEFSETPSPDGGSNHVTYSRVGVCVNDAGGQRYLVNKMSTFLKARLVCSVP 240
Db 241 DKVYFEFSETPSPDGGSNHVTYSRVGVCVNDAGGQRYLVNKMSTFLKARLVCSVP 393
QY 394 GAETHFQDLEDFVLLMPKAGKSLVYALFSTVSANVPOGAVCYHMAIDWEFNGPFAIR 453
Db 301 GAETHFQDLEDFVLLMPKAGKSLVYALFSTVSANVPOGAVCYHMAIDWEFNGPFAIR 453
QY 454 DGPQHOMGPRYGVKVPFRPGVCSKMTAOPGRPFSTKDYPRDEVLOFARHPLMFWPVR 513
Db 361 DGPQHOMGPRYGVKVPFRPGVCSKMTAOPGRPFSTKDYPRDEVLOFARHPLMFWPVR 513
QY 514 RHGRPVLYKTHLAQQLHQLIVDRAEDGTVDYIFLGTDSGSYLKVYALQAGSAPREY 420
Db 421 RHGRPVLYKTHLAQQLHQLIVDRAEDGTVDYIFLGTDSGSYLKVYALQAGSAPREY 573
QY 574 VLEELQVRFVPTITMEISVKRMLYVGSRLGYAQLRLHOCETTYGACAECCCLARDPYC 480
Db 481 VLEELQVRFVPTITMEISVKRMLYVGSRLGYAQLRLHOCETTYGACAECCCLARDPYC 633
QY 634 AMDASCTHYRPSLGRKRRFRRODIRHGNPALQCLGOSQEBEAVGLVAATMYGTENSTF 693
Db 541 AMDASCTHYRPSLGRKRRFRRODIRHGNPALQCLGOSQEBEAVGLVAATMYGTENSTF 693
QY 694 LECTPKSPQAAVAMLLQRRGDEGPDQVKTDERVLTENKLLRRLSRDAGTYCTTLEH 600
Db 601 LECTPKSPQAAVAMLLQRRGDEGPDQVKTDERVLTENKLLRRLSRDAGTYCTTLEH 753
QY 754 GFSONYVRLALVYIVASQDLNLFPEPKREBPARGCLASTPPRMYKDIQLIGFANLP 660
Db 661 GFSONYVRLALVYIVASQDLNLFPEPKREBPARGCLASTPPRMYKDIQLIGFANLP 813
QY 814 RVDYECERVMKRGTEGCGCRRSRGKQANGKSMAGLELKKKSVHAEHNRTPREVE 720
Db 721 RVDYECERVMKRGTEGCGCRRSRGKQANGKSMAGLELKKKSVHAEHNRTPREVE 873
QY 874 AT 875
Db 781 AT 782

```

```

PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167822.
PR 15-FEB-2000; 2000US-0298733.
PA (ALPH-) ALPHAGEN INC.
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
PR WPI: 2000-549267/50.
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
PS as well as nutritional sources or supplements -
PS Disclosure: Page 298-300; 309pp; English.
CC The present invention is concerned with a number of secreted proteins
CC and their coding sequences isolated from various human cDNA libraries.
CC The proteins and coding sequences can be used in the isolation of
CC similar genes and proteins, in the elucidation of their function of
CC and to treat a number of conditions. It is possible that they may have
CC factors, in immune modulation, where they may be used to treat immune
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity. No information about sequences
CC AB23632-B23645 is given in the specification.
CC XX
CC XX
Sequence 785 AA:
Query Match 88.6%; Score 4203.5; DB 21; Length 785;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 782; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
QY 94 MAPSAMALCWLIGGLLHGGSSGSPSPVRLRLSYRDLISANRAIFLPGQSLNQA 153
Db 1 MAPSAMALCWLIGGLLHGGSSGSPSPVRLRLSYRDLISANRAIFLPGQSLNQA 60
QY 154 MYLDEYDRFLFLGDLALYSILRLDQAMPDPREYLPPOGREGCRKGRDPLTECANFV 213
Db 61 MYLDEYDRFLFLGDLALYSILRLDQAMPDPREYLPPOGREGCRKGRDPLTECANFV 120
QY 214 RVLPNHNTHLACGTGAFOTCALITVGHREHVLHLEPGSVESGRGRCHEPSPFAS 273
Db 121 RVLPNHNTHLACGTGAFOTCALITVGHREHVLHLEPGSVESGRGRCHEPSPFAS 180
QY 274 TFIIDGLYGLTADFLGREA--MIFRSGGPRPALRSDDSLHDPFVMAARIPENS 330
Db 181 TFIIDGLYGLTADFLGREAMIFRSGGPRPALRSDDSLHDPFVMAARIPENS 330
QY 331 QDNKVVYFSEETVSPDGGSNHVTYSRVGVCVNDAGGQRYLVNKMSTFLKARLVCSVP 300
Db 241 QDNKVVYFSEETVSPDGGSNHVTYSRVGVCVNDAGGQRYLVNKMSTFLKARLVCSVP 390
QY 391 GPGAETHFQDLEDFVLLMPKAGKSLVYALFSTVSANVPOGAVCYHMAIDWEFNGPFA 450
Db 301 GPGAETHFQDLEDFVLLMPKAGKSLVYALFSTVSANVPOGAVCYHMAIDWEFNGPFA 450
QY 451 AHRDGPQHOMGPRYGVKVPFRPGVCSKMTAOPGRPFSTKDYPRDEVLOFARHPLMFWPVR 513
Db 361 AHRDGPQHOMGPRYGVKVPFRPGVCSKMTAOPGRPFSTKDYPRDEVLOFARHPLMFWPVR 513
QY 513 VBRHGRPVLYKTHLAQQLHQLIVDRAEDGTVDYIFLGTDSGSYLKVYALQAGSAPREY 573
Db 421 VBRHGRPVLYKTHLAQQLHQLIVDRAEDGTVDYIFLGTDSGSYLKVYALQAGSAPREY 573
QY 573 BEVYLELQVRFVPTITMEISVKRMLYVGSRLGYAQLRLHOCETTYGACAECCCLARD 630
Db 481 BEVYLELQVRFVPTITMEISVKRMLYVGSRLGYAQLRLHOCETTYGACAECCCLARD 630
QY 631 PYCAMDASCTHYRPSLGRKRRFRRODIRHGNPALQCLGOSQEBEAVGLVAATMYGTENH 690

```


|||||
 541 PYAMGASCTHRRSLGRFRFRDINHGNPLOCIGOSOEFAVAVATVSTENH 600
 691 STELECPSPQAAVWILORPGDEGPDQVKTDERVLTFFENGILLFRILSPFDAGTYCTT 750
 601 STELECPSPQAAVWILORPGDEGPDQVKTDERVLTFFENGILLFRILSRDAGTYCTT 660
 751 LEHGEQTVVRLALVIVASOLDNFFPPPEPPPPARGLASTPPKAWYDIIOLIGFA 810
 661 LEHGEQTVVRLALVIVASOLDNFFPPPEPPPPARGLASTPPKAWYDIIOLIGFA 870
 811 NLPVDECEVWRCGTETSCGCFRFRSRSRCKOAKGKSMAGLELGGKMKSRVHAENRTPR 870
 721 NLPVDECEVWRCGTETSCGCFRFRSRSRCKOAKGKSMAGLELGGKMKSRVHAENRTPR 780
 871 EWEAT 875
 781 EWEAT 785

RESULT 5
 AAB84219 standard; protein; 779 AA.

06-AUG-2001 (first entry)

Amino acid sequence of a semaphorin polypeptide designated ZSMF-16.
 Human: semaphorin; ZSMF-16; neurite growth; neurite outgrowth;
 T lymphocyte suppressor; cancer; stroke; brain damage; paralysis;
 spinal injury; neurodegenerative disease; amyotrophic lateral sclerosis;
 Alzheimer's disease; Huntington's disease; Parkinson's disease;
 peripheral neuropathy; demyelinating disease; multiple sclerosis;
 immunosuppression; autoimmune disease; insulin dependent diabetes;
 rheumatoid arthritis.

Location/Qualifiers
 Key 1..22 "signal peptide"
 Peptide /note="signal peptide"
 Protein 23..779
 Modified-site /note="mature protein"
 Domain 62
 /note="N-glycosylation site"
 /note="76..500
 /note="semaphorin"
 Modified-site 124
 /note="N-glycosylation site"
 Domain 593..654
 /note="Ig-like domain"
 Modified-site 594
 /note="N-glycosylation site"

MO200140278-A2.
 07-JUN-2001.
 06-DEC-2000: 2000MO-US33116.
 06-DEC-1999: 99US-0455560.
 (ZYMO) ZYMOGENETICS INC.
 Holloway JL, Foley KP;
 WPI: 2001-374784/39.
 N-PSDB: AAF90250.
 Novel human semaphorin polypeptide, ZSMF-16, useful for treating
 peripheral neuropathies Alzheimer's and Huntington's disease and

polynucleotide encoding ZSMF-16 useful for detecting genetic
 abnormality and cancer

Claim 13; Page 121-123; 124pp; English.
 The present sequence represents a semaphorin polypeptide, designated
 ZSMF-16. ZSMF-16 is a neurite growth and development modulator. It also
 enhances spinal cord and sensory neurite outgrowth and patterning, and
 is involved in the activation and regulation of T lymphocytes suppressor.
 ZSMF-16 is useful for detecting a genetic abnormality or cancer. ZSMF-16
 polynucleotide probes can be used to detect 3p21 loss, trisomy,
 duplication or translocation associated with cancer, etc.. ZSMF-16 can
 be used to modulate neurite growth and development and demarcate nervous
 system structures. ZSMF-16 are also useful for regenerating and directing
 neurite outgrowth following strokes, brain damage caused by head
 injuries, paralysis caused by spinal injuries, and for treating
 neurodegenerative diseases such as amyotrophic lateral sclerosis,
 Alzheimer's disease, Huntington's disease, Parkinson's disease and
 peripheral neuropathies, or demyelinating diseases e.g., multiple
 sclerosis. ZSMF-16 also acts as a mediator of immunosuppression, such as
 insulin dependent diabetes, rheumatoid arthritis, and multiple
 sclerosis. It can also be used as an anti-inflammatory for inhibition
 of antigen in humoral and cellular immunity and for immunosuppression
 in graft and organ transplants.

Sequence 779 AA:
 Query Match 85.3%; Score 4046.5; DB 22; Length 779;
 Best Local Similarity 95.0%; P-Val 0; Indels 39; Gaps 2;
 Matches 760; Conservative 0; Mismatches 115 135

94 WAPSAKALCMLLGLLHGGSSGSPGSPVPRRLSTYD-----ILS 135
 1 WAPSAKALCMLLGLLHGGSSGSPGSPVPRRLSTYD-----ILS 135
 136 ANRSALFLGPGSLTQWYLYDEYRDLFLGLDALYSLRLDQAWDPREVLPQPGQR 195
 61 ANRSALFLGPGSLTQWYLYDEYRDLFLGLDALYSLRLDQAWDPREVLPQPGQR 195
 196 EECVRRGRDPLTECANFVRLDPHNRTHLACGTGAFOPTCALITVGHREHYHLEPGS 159
 111 -----TECANFVRLDPHNRTHLACGTGAFOPTCALITVGHREHYHLEPGS 159
 256 VESGRGCPHEPRSPASTFIDELTYGLTADPLGREAMIFRSGGPRPALRSDOSLH 315
 160 VESGRGCPHEPRSPASTFIDELTYGLTADPLGREAMIFRSGGPRPALRSDOSLH 315
 316 DREFVMAARIPENDNDKVFEEFSETVSPDGSNHYTVSRVGVYCNVAGGQRLVN 279
 220 DREFVMAARIPENDNDKVFEEFSETVSPDGSNHYTVSRVGVYCNVAGGQRLVN 279
 376 KWSTFLKARLYCSVPGGAEETHDQLEDFLLMPKAGKSELYVALFSTYSAVFOGFANC 435
 280 KWSTFLKARLYCSVPGGAEETHDQLEDFLLMPKAGKSELYVALFSTYSAVFOGFANC 435
 436 VYHMAADIMEVFNCPFAHRDGPQHOMGPYGGKYPFRPGVCSKMTAOPGRFGSTKDPD 495
 340 VYHMAADIMEVFNCPFAHRDGPQHOMGPYGGKYPFRPGVCSKMTAOPGRFGSTKDPD 495
 496 EVLOFARAHPLMPVPRRGRPRVLTAKTHLAQDLHIVYDRVADGDTYDVFILGTDSGS 555
 400 EVLOFARAHPLMPVPRRGRPRVLTAKTHLAQDLHIVYDRVADGDTYDVFILGTDSGS 555
 556 VLKVALALOGGSAPEEVEVLELOVFKYPPITMEISYVRKMLYVGSRLGVNOLRLHQC 615
 460 VLKVALALOGGSAPEEVEVLELOVFKYPPITMEISYVRKMLYVGSRLGVNOLRLHQC 615
 616 EYGTACACCCCLARDPYCAMDGASCTHYRPSLGRKRRRRODIRGNPALOCIGOSOEFA 579
 520 EYGTACACCCCLARDPYCAMDGASCTHYRPSLGRKRRRRODIRGNPALOCIGOSOEFA 579

OY 676 VGLVAATVYCTEHNSTLECLPKSPQAAVRLQRPDGPDKYKTDERVLHTEGGLF 735
 Db 580 VGLVAATVYCTEHNSTLECLPKSPQAAVRLQRPDGPDKYKTDERVLHTEGGLF 735
 OY 736 RLRSFAGTYCTTLEHFSQTVVRLALVVASQDNLFPPEPKPEEPARGGLASTP 795
 Db 640 RLRSFAGTYCTTLEHFSQTVVRLALVVASQDNLFPPEPKPEEPARGGLASTP 795
 OY 796 PKAWYDIIQILGFANLPRVDEYCEKWCRCGTEGSCGFSRSRKGQARKSMAGLELCK 855
 Db 700 PKAWYDIIQILGFANLPRVDEYCEKWCRCGTEGSCGFSRSRKGQARKSMAGLELCK 855
 OY 856 KKKSVNAEHNTREVEAT 875
 Db 760 KKKSVNAEHNTREVEAT 875

RESULT 6

AAM63748
 ID AAM63748 standard; Protein; 775 AA.
 AC AAM63748:

DT 01-OCT-1998 (first entry)
 XX

DE Human semaphorin.
 XX

KW Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
 KM neurological disease; atopic skin inflammation; autoimmune disease;
 XX pain.
 OS Homo sapiens.

PN MO9822504-A1.
 PD 28-MAY-1998.
 XX

PF 12-NOV-1997; 97WO-JP04111.
 PR 15-NOV-1996; 96JP-0321068.
 XX

PA (SDMU) SUMITOMO PHARM CO LTD.
 PI Furuyama T, Inagaki S.
 DR WPI; 1998-312416/27.
 DR N-PSDB; AAV35367.

PT Gene encoding new semaphorin nerve growth inhibitor - useful in
 XX diagnosis, treatment and study of neurological diseases
 PS Claim 1; Page 33-37; 49pp; Japanese.

CC The present sequence represents human semaphorin, a nerve growth
 CC inhibitor. The semaphorin protein, and gene encoding the protein,
 CC and their derivatives, are used in the diagnosis, treatment and
 CC study of neurological disorders such as atopic skin inflammation,
 CC autoimmune diseases and pain.
 SO Sequence 775 AA.

Query Match 43.2%; Score 2052.5; DB 19; Length 775;
 Best Local Similarity 50.7%; Pred. No. 8.9e-177;
 Matches 399; Conservative 127; Mismatches 234; Indels 27; Gaps 10;

OY 94 MASAAMICWILGILLHGGSSGSPGSPVRLRLSVRDLLSARSAIFLGPOGSLNLOA 153
 Db 1 MAVAGHITLLMLGLHLEWTFGHSANSPYRLRLSHKELLELRKTSIFOSPLGFLDHT 60
 OY 154 MYLDEYRDLEFLGIDALYSRLDQAMPDPREVLMPPQPGQRECVARKGRDPLTECANL 213
 Db 61 MLDEYOEKLELVGGADLVYSLNLERVSDGYREIYWSTAVKVECIKMGKD-ANECANI 119

OY 214 RVLQPHNTHLLACGTAFQPTCALITYGHRSHV-LHEPGSVESGRCRCPHEPSRFA 272
 Db 120 RVLHNYNTHLLCATGAFDPHCATIRVGHSHSEPLFHLHSHSRSGRCRCPHEPSRFA 272
 OY 273 STFDIGELVYGLADFLGREAMIFRSRGRPALRSD-IDOSLHDPREPVMAARIPEMSDQ 331
 Db 180 STVAGNELFAGVLDYWGDSALFRSMKLGHIIFHEHDEMLKPEFVSGVMIPEMDR 239
 OY 332 DNDKVFEESEIYPPSDGSGNHVTVSRVGYCNDAGORVLYKWSSTFLKARLVCSVG 391
 Db 240 DDKRYEFTTEKALAEENNA-HTTYTRVGRCLVDMGQRLVANKSTFLKARLVCSVG 298
 OY 392 PGAEHEDQLEDFVFLMPKAGKSLVYALFSTYSAFVGAVCYHMDIIEVEENGPA 451
 Db 299 AMGIDYFDELEDFELLPTRPKNPVIFGLFNTTSIFRGHVCYHMSIIEARNGPYA 358
 OY 452 HRDGFQHMGPYGGKVPFPPRGVCPKMTAOPGRPGSTKDYPPDEVLOFARHPLMEWPV 511
 Db 359 HKEGPEYHWSLYEGKVYPPRGSCASKVN--GKYGTTKDYPPDAIRFARHPLMOPV 415
 OY 512 RPRHGRPVLVKTHLAQQLQIYVRYEAEDGTVDYIFLGTSGSVLYALQAGSAPE 571
 Db 416 KPVHKKPILVKTGDKYNLQALVDREAEADQYDIFLIGDTGLVAVITYNOETEMME 475
 OY 572 EYVLELOVFKVPTPTTEMTISVKRMLYGSRLGVQALRHOCETVGTACACCLARP 631
 Db 476 EYVLELOVFKVPTPTTEMTISVKRMLYGSRLGVQALRHOCETVGTACACCLARP 631
 OY 632 YCAMDASCCTHYRP--SLGKRPRRODIRHGNPALQCLGSOEEAVGVLAATVYGTEN 689
 Db 536 YCAMDASCCTHYRP--SLGKRPRRODIRHGNPALQCLGSOEEAVGVLAATVYGTEN 689
 OY 690 NSTFLECLPKSPQAAVRLQRPDGPDKYKTDERVLHTEGGLFRLRSFADAGTYCTT 749
 Db 596 NSTFLECLPKSPQAAVRLQRPDGPDKYKTDERVLHTEGGLFRLRSFADAGTYCTT 749
 OY 750 TIEHFSQTVVRLALVVASQDNLFPPEPKPEEPARGGLASTP--WYKEE 713
 Db 804 TIEHFSQTVVRLALVVASQDNLFPPEPKPEEPARGGLASTP--WYKEE 713
 OY 714 LQILGFANLPRVDEYCEKWCRCGTEGSCGFSRSRKGQARKSMAGLELCKMKSRVHA 863
 Db 864 LQILGFANLPRVDEYCEKWCRCGTEGSCGFSRSRKGQARKSMAGLELCKMKSRVHA 863
 OY 864 EHNKTPR 870
 Db 764 EHNKTPR 870

RESULT 7

AAY43090
 ID AAY43090 standard; Protein; 775 AA.
 AC AAY43090:

DT 05-JAN-2000 (first entry)
 XX

DE Mouse semaphorin H (Sema H) amino acid sequence.
 KW Semaphorin H; Sema H; Sema HV; collapsin; bone structure formation;
 KM metastasis; cancer; antibody; drug screen.
 XX Mus sp.

FH Key Location/Qualifiers
 FT Misc-difference 615
 FT /note="Encoded by GTA"
 XX

PN MO9947671-A2.
 PD 23-SEP-1999.
 PF 12-MAR-1999; 99WO-IB00495.

CC metastasis. The invention uses the mouse *Sema H* gene in the diagnosis of
 CC metastatic cancer. Semaphorin polynucleotides and polypeptides are used
 CC in the methods of the invention, the polypeptides and polypeptides are used
 CC determining the metastatic potential of cells, by detecting their
 CC expression in biological samples. Antibodies specific for *Sema H*, are
 CC also useful therapeutically in inhibiting *Sema H* polypeptide activity and
 CC polypeptide using e.g. a small molecule inhibitor or a semaphorin-H
 CC ligand (or fragment). The polynucleotides can also be used to inhibit
 CC prevent metastasis of cancer cells using known antisense technology e.g. to
 CC quantify *Sema H* mRNA levels in cells. The polypeptides, fusion proteins,
 CC multimeric proteins, antibodies or antisense oligonucleotides can be
 CC included in pharmaceutical compositions. The polynucleotides can be used
 CC to isolate similar sequences from other species and to produce mammalian
 CC cell lines and tumours with known metastatic potential, useful in
 CC anti-metastatic drug screening.

Sequence 777 AA:

Query Match 40.98; Score 1942.5; DB 20; Length 777;
 Best Local Similarity 49.08; Pred. No. 8.5e-167;
 Matches 387; Conservative 125; Mismatches 248; Indels 29; Gaps 12;

94 MAPSAMALCMLGGLLHGSSGSPSPVRLRLSYRDLANSATFLGPOSLNQA 153
 1 MAPAGHILTLMLGHLLEMTWPGHSANPSYARLPISHKEFLNGLOFYKAPLGLDHT 60
 154 MYDEYDRFLFLGGLDALSLRDAMPDPREVLPPOGORECVKGGDPLTECANEF 213
 61 MLDEYQERLFLGGRDLYSLNLERVSDGYREIYWPSTAVKVECIKMGD-ANECANYI 119
 214 RVLOPHNRTHLLACGTGAFPTCALITVGRHGEHL-ILPBGVESGRGRCHPERSRPA 272
 120 RVLIHYNRTHLLCATGAFPHCAFIVGHNSEBPLFHLSEHSEGRGRCHPERSRPA 272
 273 STFDIGELTGLTADLTGREGAMIFRSGCRPRALRSD-SQSLHDPREVMARIPENDQ 331
 180 STVGNELFLGGLSDWGRSALFISMGRLGHTEDHDERLKEPKFVGSMYLPDNEBR 239
 332 DNDKVFSESEIVSPDGGSNHVTASRVCVNDAGORLVNKMSTFLKARLVCSVP 391
 240 DDKMTFFETKALEENNA-HTILHPSGRICLVNDMGORITLVKMTFLKARLVCSVP 298
 392 PGCAETHFDQLEDFELMPKAGKSLLEYVALFSTVSAVFGFACVYCHMDIWEFNGRPA 451
 299 MNGIDTYDELEDFELFTRPKNPVIFGLNTSNIFRGHACVYHMSISREAFNGRPA 358
 452 HRDPOHMGVYGGKVRPRRVCSPSKMTAOPRPRGSKDDYDEVLOFARAHPLMPRV 511
 359 HKEBPEYHMSLEYKVPYRPRSCASKVN---GKKGITQRLRDLAIFRAPHNPLMQEI 415
 512 RPRGRVULKTHLAQQLQIYVDRVEADGTUUYIFIGTDSGSV-LKVIALDAGSGAP 570
 416 KPVHKKRPIVLKTDGKYNLQALVDRVEADGUYLFIGTDTGTLVLTITNOETEMW 475
 571 EEVULLELOVFPVPTITEMEISYRKMLYGSLGVAOLRKHOCSEYTGACAECLAND 630
 476 EEVULLELOVFPVPTITEMEISYRKMLYGSLGVAOLRKHOCSEYTGACAECLAND 630
 631 RYCDAGSASTHNPSSG--KRRFRDRIHGNAPALOCGOSOEBAAGLVATVNGTE 688
 536 RYCDAGSASTHNPSSG--KRRFRDRIHGNAPALOCGOSOEBAAGLVATVNGTE 688
 689 HNSTFLFELKPSQAAVNRVLORPDGDEPDQYKTERVILNTERGLFRLSRFAGIYTC 595
 596 SNSTLLECPRLSLQAKVIVFQKGRDVRKEEVKTDKRVKMKDLGLLRLVAKSDAGIYTC 655
 749 TTEHEFSTVRLALVIVASQDLNLFRRPRKPE-----PFRAGLSTPRKMYKD 802
 656 QVNEINFTVHTAKITLLEVVEHKEGMEFKNDHEERHNKMRCPRLSMOGSTPR--WYKE 713

OY 803 ILQILGFAN-LPRVDEYCEYWCRTTECSGCFRSRSGKQARGNAGLELGRKMSRV 861
 DB 714 FLQILGSGFQVEYERCEWCTD-----KRRKTKMSPSKMYANPDEKRLRSK- 764
 OY 862 HAHNRTPR 870
 DB 765 -AEHFRPLR 772

RESULT 9

AAG62730

ID AAG62730 standard; peptide: 777 AA.

AC AAG62730;

DT 17-SEP-2001 (first entry)

DE Amino acid sequence of mouse semaphorin *Sema3E*.

KW Hanatoxin; tarantula; toxin; voltage-gated potassium channel;

KW voltage-gated calcium channel; hanatoxin-like sequence; HTLS;

OS semaphorin; dorsal root ganglion repulsion; growth cone collapse.

Mus sp.

WO200138491-A2.

31-MAY-2001.

PF 07-NOV-2000; 2000WO-US41943.

PR 08-NOV-1999; 99US-0164056.

PA (GENO) GEN HOSPITAL CORP.

PI Behar O, Woolf CJ;

DR WPI; 2001-451494/48.

PT Polypeptide sequences that encompass the hanatoxin-like sequences of
 PT semaphorins, useful as drugs to treat any condition or disease that
 PT is characterized by abnormal calcium channel function.

PS Claim 6; Page 13; 29pp; English.

CC The present sequence represents a semaphorin. The specification
 CC describes hanatoxin-like sequences (HTLS) found in the semaphorin
 CC domain of mammalian secreted semaphorins. Hanatoxin is a tarantula
 CC toxin that selectively blocks some voltage-gated potassium and calcium
 CC channels. The HTLS is responsible for the dorsal root ganglion repulsion
 CC and growth cone collapse activities associated with semaphorins.
 CC Polypeptides containing HTLS can be used to modulate the activity of
 CC calcium channels. The peptides can also be used as an antigen to
 CC generate antibodies that can then be used to modulate the activity of
 CC the peptides or antibodies can be used as drugs to treat any condition
 CC or disease that is characterized by abnormal calcium channel function.

Sequence 777 AA:

Query Match 40.98; Score 1942.5; DB 22; Length 777;
 Best Local Similarity 49.08; Pred. No. 8.5e-167;
 Matches 387; Conservative 125; Mismatches 248; Indels 29; Gaps 12;

94 MAPSAMALCMLGGLLHGSSGSPSPVRLRLSYRDLANSATFLGPOSLNQA 153
 1 MAPAGHILTLMLGHLLEMTWPGHSANPSYARLPISHKEFLNGLOFYKAPLGLDHT 60

154 MYDEYDRFLFLGGLDALSLRDAMPDPREVLPPOGORECVKGGDPLTECANEF 213
 61 MLDEYQERLFLGGRDLYSLNLERVSDGYREIYWPSTAVKVECIKMGD-ANECANYI 119

214 RVLOPHNRTHLLACGTGAFPTCALITVGRHGEHL-ILPBGVESGRGRCHPERSRPA 272

Db 591 SSTELESPKSRQALVYVWQFQRNEERKEIRVDHIIIRDOGLLRSLOQDSGNVLTCH 650
 QY 750 TLEHGFQSVVRLALVYIVASQDLNLFPEPEKPEPPAGGLASTPP--KAWYKDLLOI 807
 Db 651 AVEHGFQITLLKLVLEIYDHELELLHKDDGSGSKTKEMSNMTPSOQKVVYRDFMQLI 710
 QY 808 GFANLPRDVEYCEWRCGTTECSGCFRSRSGKARGS--WAGLEIGKKMKSRVNAEH 865
 Db 711 NHPNLNTWDECEQVWKRDRKQ-----RROPRGHTPGNSNMKMKHLOENKGRNRRTHEF 764
 QY 866 NRPREV 872
 Db 765 ERAPRSV 771

RESULT 11
 AAG62726
 ID AAG62726 standard; peptide; 771 AA.
 AC AAG62726;

17-SEP-2001 (first entry)

Amino acid sequence of human semaphorin Sema3A.

Hanatoxin; tatanula; toxin; voltage-gated potassium channel;
 voltage-gated calcium channel; hanatoxin-like sequence; HTLS;

semaphorin; dorsal root ganglion repulsion; growth cone collapse.

Homo sapiens.

MO200138491-A2.

31-MAY-2001.

07-NOV-2000; 2000MO-US41943.

08-NOV-1999; 99US-0164056.

(GEHO) GEN HOSPITAL CORP.

Behar O, Woolf CJ;

WPI: 2001-451494/48.

Polypeptide sequences that encompass the hanatoxin-like sequences of
 semaphorins, useful as a drugs to treat any condition or disease that
 is characterized by abnormal calcium channel function.

Claim 6; Page 11; 29pp; English.

The present sequence represents a semaphorin. The specification
 describes hanatoxin-like sequences (HTLS) found in the semaphorin
 toxin that selectively blocks some voltage-gated potassium and calcium
 channels. The HTLS is responsible for the dorsal root ganglion repulsion
 and growth cone collapse activities associated with semaphorins.
 Polypeptides containing HTLS can be used to modulate the activity of
 calcium channels. The peptides can also be used as an antigen to
 generate antibodies that can then be used to modulate the activity of
 calcium channels by inactivating naturally occurring channel ligands.
 The peptides or antibodies can be used as drugs to treat any condition
 or disease that is characterized by abnormal calcium channel function.

Sequence 771 AA;

Query Match

Best Local Similarity 38.6%; Score 1830; DB 22; Length 771;

Matches 366; Conservative 139; Mismatches 244; Indels 38; Gaps 13;

101 ICWLLGLLHGSSSPSPGSPVPRLLSYRDLISANRSARITLGGSLNLTQAMVLEVR 160

Db 8 VC-LFNGVLTARANYONGKNVPRLLSYKEMLESNNVITENGIANSSVHTLELDEER 66
 QY 161 DRPLGLALVALYLRLQDAMPREVLMPPOGQDEECVGRKGRDPLTECAFEVAVLQPHN 220
 Db 67 SRLVYAKAHIFSFDLVNI-KDFOKIWPVSTRDECKRMAGKIDILKECANFIVLKAVN 125
 QY 221 KPHLLACGTAGAPFCALITVCHGE-HVLLHEGVSVEGRCRPHRSPRSTFIDGE 279
 Db 126 QTHLYACGTGAHFIRCTEITIGHHEDNIEKLENSHENGRCGSPYDKLLSLIDGE 185
 QY 280 IYTGTLADELGRAMIRFGSGPRELRSDS-DOSLHDPFVMAATIPENSODNDKVF 338
 Db 186 LYSGLADPFGMDALFRITGLHHPIRTEQHDHSRWLNDREFISAHILISSDNPEDDKYF 245
 QY 339 FFESEVSPDG-GSNHVTASRVCVNDAGQRYLVNKNSTFLNARLVCSVPGGAE 397
 Db 246 FFREN-AIDEGHSGKATHRIQICKNDPFGHRLSVNKTFLKARLISVPGNGIDT 303
 QY 398 HFDOLEDFVLLMPKAKSLLEVYALSTVSAPVQGFVAVCYVMAADIMEVFNPAHRDGP 457
 Db 304 HFDELQDFLNMFKDKKNVYVGVFTTSNIEKGSAYCMGSDYKRVFLGPAVHRDGP 363
 QY 458 HMGPRYGVKVPFPPRQVCSKMTAOPGRPGSTKDYPRVLOFPAHPLMFWPRPHGR 517
 Db 364 YQWVPYQGVNVPYPRGTGCPSTKTFG-GFDSTDLDDVITFARSHPMYNPFPNNR 419
 QY 518 PVLVYTHLAQQLHQLVDRVDAEDGYDIFLGTDSGSVLRVATLQAGSAPREPVLEE 577
 Db 420 PLVKTQVNYQFTQIVDRVDAEGQYDMFTCTDGTVLKVSIPKRTWYDELEEE 479
 QY 578 LQVFPVPTFEMETISVRKMLYVGSRLGYAQLRHQCEYVGAACGLADPYCAMDG 637
 Db 480 MTFVREPTAISAMELSKQOOLYIGTAGVAOLPHRCDIYKACABECLADPYCAMDG 539
 QY 638 ASCTHYRPSLGRFRRODIRNGPALQCL-----QOSEEAVGLVATVYGTEN 689
 Db 540 SACSRYFPT-ARRTRRQDIRNGDPLTHCSDLHNDHNSHPER-----IIVGVN 590
 QY 690 NSTFLECLPRKSPQAAVWMLLOPREGDQYKTDRLVHTERGLFRRLSPAGCTYCT 749
 Db 591 SSTELESPKSRQALVYVWQFQRNEERKEIRVDHIIIRDOGLLRSLOQDSGNVLTCH 650
 QY 750 TLEHGFQSVVRLALVYIVASQDLNLFPEPEKPEPPAGGLASTPP--KAWYKDLLOI 807
 Db 651 AVEHGFQITLLKLVLEIYDHELELLHKDDGSGSKTKEMSNMTPSOQKVVYRDFMQLI 710
 QY 808 GFANLPRDVEYCEWRCGTTECSGCFRSRSGKARGS--WAGLEIGKKMKSRVNAEH 865
 Db 711 NHPNLNTWDECEQVWKRDRKQ-----RROPRGHTPGNSNMKMKHLOENKGRNRRTHEF 764
 QY 866 NRPREV 872
 Db 765 ERAPRSV 771

RESULT 12

AAY21264 standard; Protein; 796 AA.

AAY21264;

22-JUL-1999 (first entry)

Human semaphorin III wild type protein fragment 1.

Human; beta-amyloid precursor protein; beta-Ap4; diagnosis; cancer;
 Alzheimer's disease; age-related disease; neurodegenerative disorder;
 Huntington's disease; Down's syndrome; myotonic dystrophy; neuronal;
 diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 neurofilament-F; presenilin 1; presenilin 2; cellular tumour antigen;
 gliol fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;

bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene. HMGP-C; NSP-A;
 KM high mobility group protein-C; neuroendocrine specific protein A.
 XX Homo sapiens.
 XX M09845322-A2.
 XX 15-OCT-1998.
 XX 02-APR-1998; 98MO-1B00705.
 XX 10-APR-1997; 97US-0043163.
 XX (UYUT-) RIJKSUNIV UTRECHT.
 XX (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
 XX (UYRO-) UNIV ROTTERDAM ERASMUS.
 XX Burbach JPH, Grosveld FG, Van Leeuwen FW.
 XX WPI: 1998-609901/51.
 XX N-PSDB: AAX575767.
 XX Diagnosing disease by detecting frameshift mutations in RNA or
 XX corresponding protein mutations - used to diagnose cancer and
 XX neurological diseases, particularly Alzheimer's disease, and also
 XX for treatment and prevention with specific ribozymes or wild-type
 XX RNA
 XX Disclosure: Figure 16; 258bp; English.
 XX This invention describes a novel method for the diagnosis of a disease
 XX caused by, or associated with, an RNA molecule that has a frameshift
 XX mutation. The method is used to diagnose age-related diseases, especially
 XX cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
 XX disease, Down's syndrome, myotonic dystrophy, Huntington's disease, II
 XX multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
 XX and many others listed) or susceptibility to these disorders. The method
 XX allows a definitive diagnosis of Alzheimer's disease in living patients,
 XX at an early stage. It is based on the observation that disease may be
 XX caused by mutations in RNA rather than DNA. The invention describes the
 XX use of neuronal system RNA molecules, specifically proteins including
 XX beta-amyloid precursor protein (beta-APP), the microtubule associated
 XX protein tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
 XX associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
 XX neurofilament-E, presentin I, presentin II, glial fibrillary acidic
 XX protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
 XX 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
 XX protein-C (HMGP-C) and neuroendocrine specific protein A.
 XX Sequence 796 AA:
 XX 38.1%; Score 1810; DB 19; Length 796;
 XX Query Match 45.4%; Pred. NO. 9.2e-155; Indels 58; Gaps 14;
 XX Best Local Similarity 139; Mismatches 244;
 XX Matches 366; Conservative 139; Mismatches 244;
 XX 101 ICMILGILLHGSGSPSPSPRLSYRLDLSANSATFLGPGSLNQAAYIDEVR 71
 XX 13 VC-LFMGVLLTARANYQNGKNNVPRKLKSYKELSNVYIFNGLSNSSYHTFLDEBR 71
 XX 161 DRFLPGDALYSRLDOAMPDPREVILMPPOGQREECYRKGRDPLETCANFVRLPHN 220
 XX 72 SFLYVGADHIFESFDLVNI-KDFOKIWPVSYTRDCKKAGDLKECANFKIKLAKVN 220
 XX 221 RTHLACGTGAFOPTCALITVGHGRGE-HVLLLEPGSVESGRGCPHEPSRPFASFIDE 279
 XX 131 QTHLYNCGTGAHPICTYIEIGHNEDNIFLNSHFENGRGSPYDKLITSLIDEE 190
 XX 280 LYTGLADELGRAMTFRSRGPRALRSDS-DOSLHDPFVMAARIENSDODNKKYF 338
 XX 191 LYSTADAEFGWGRDFAIFRLGHHHPRTEQDHSRWLNPKFISAHLSIESDNPDDVYF 250
 XX 339 FFSFTVSPDG-GSNHVTYSRVGVCVNDAGORVLVKNWSPFLKARLVCSVPGGAET 397

251 FERN--AIDGESHKATHARIGOI CKNDPGHNSLVNKKWTFILKARLICSVDPGPNIGDT 308
 398 HFDQLEDVFLMKRACKSLLEVAFSTVSAVFOGFAVCYHMAIDWEVFNPGFARHDPQ 457
 309 HFDQLEDVFLMKRACKSLLEVAFSTVSAVFOGFAVCYHMAIDWEVFNPGFARHDPQ 368
 458 HONGPYGKVPFPPRGVCSKMTAQRGPRGSKDYDEVLQFARAPDLMFVPR----- 513
 369 YQWVPYQGRVPRPGTCSKSTGSGDSTKDLDDYIFPARSHPMYMPVPMNKR 424
 514 -----RGRPVLVKTHLAQQLHDIYVDREADGFTYDVIPLGDSGYL 557
 425 PIVIKTDVNTQFTQVIMNNRPVIVIKTDVNTQFTQVIMNNRPVIVIKTDVNTQFTQV 484
 558 KYIALDAGSAADEVEVLELOFVKVPPITEMELSVKRMVYGSALGVAOLRLHOCET 617
 485 KYVSIPEKWTYDLEVEVLELOFVKVPPITEMELSVKRMVYGSALGVAOLRLHOCET 544
 618 YGTACACCCCLARPYCAMDGASCTHTRPSLGRRRPRRODIRHGNALOCIT-----GQ 669
 545 YKACACCCCLARPYCAMDGASCTHTRPSLGRRRPRRODIRHGNALOCIT-----GQ 603
 670 SOEBEANGVLAATVYGFENSTFLECLPKSPQAAVRLORPGEPPDYKTERVLYLT 729
 604 SPEER-----IIVGENSTFLECLPKSPQAAVRLORPGEPPDYKTERVLYLT 655
 730 EGGLEFRLSRPDCTYCTLENGFSOTVRLALVYVASQDLNLFPEPKRPEPRPARG 789
 656 DOGLLRSLQKDSGNYLCHAVEGEIOTLKVLEYIDREHELELHKDDDGSKTKE 715
 790 GLASTPR--KAWYKDIOLIGFANLPRVDEYCEYVRCGTEGSCFRSRKQARGKS 847
 716 MSNSMTPSOKWYRDMQOLINPNTIMDEFEQYWKDRQ-----RRORGCHTPGNS 769
 848 --WAGLELGMKMSRYNAEHNRPREV 872
 770 NKMKHLENNKRRNRTHERRAPRSV 796
 RESULT 13
 AAG62727 standard; peptide: 749 AA.
 AAG62727:
 17-SEP-2001 (first entry)
 Amino acid sequence of human semaphorin sema3B.
 Hanatoxin; tarantula; toxin; voltage-gated potassium channel;
 voltage-gated calcium channel; hanatoxin-like sequence; HTLS;
 semaphorin; dorsal root ganglion repulsion; growth cone collapse.
 Homo sapiens.
 MO200138491-A2.
 31-MAY-2001.
 07-NOV-2000; 2000WO-US41943.
 08-NOV-1999; 99US-0164056.
 (GENO) GEN HOSPITAL CORP.
 Behar O, Woolf CJ;
 WPI: 2001-451494/48.
 polypeptide sequences that encompass the hanatoxin-like sequences of
 semaphorins, useful as a drugs to treat any condition or disease that
 is characterized by abnormal calcium channel function -

PS Claim 6; Page 12; 29pp; English.

CC The present sequence represents a semaphorin. The specification
CC describes hanatoxin-like sequences (HTLS) found in the semaphorin
CC domain of mammalian secreted semaphorins. Hanatoxin is a tatanula
CC toxin that selectively blocks some voltage-gated potassium and calcium
CC channels. The HTLS is responsible for the dorsal root ganglion repulsion
CC and growth cone collapse activities associated with semaphorins.
CC polypeptides containing HTLS can be used to modulate the activity of
CC calcium channels. The peptides can also be used as an antigen to
CC generate antibodies that can then be used to modulate the activity of
CC calcium channels by inactivating naturally occurring channel ligands.
CC The peptides or antibodies can be used as drugs to treat any condition
CC or disease that is characterized by abnormal calcium channel function.
XX

SO Sequence 749 AA;

Query Match
Best Local Similarity 36.8%; Score 1747; DB 22; Length 749;
Matches 367; Conservative 107; Mismatches 239; Indels 66; Gaps 17;

DB 83 GRORCPQFSPMSAMALCWLGLLHGGSSGSPSPSVRLRLSYRDL-----L 134
DB 2 GRGAAAVIPGLA-LTMV-----GL-----GSAFSP-----PRLSLFQELQAMHGLQTF 47
OY 135 SANRSAIFIGPGSGLQAMYLDEYRDLTGLDALYSLRLDQAMPREVLPMPQPGQ 194
DB 48 SLERTCCY-----QALLVDEERGRLVGAENHVASLMDNISRAKRLAMPFVEM 98
OY 195 REECYRKGRLTECANFVRLQPNRTHLACGAFOPCALITIGHNGEN-VLHLER 253
DB 99 RECMWAKKIDGTETCMNVKLLHAYNRTHLACGAFHPTCAFEVGHAEFVRLRDP 158
OY 254 GVSBSGRCPHSPSPFASFTIDELTYGLADFLGKEMIRFGSGPRRLRSD-SDOS 312
DB 159 GRIDGKSPYDRHRAASVTLGEBELYSVADLMGRDFTTIRSLGQRP-SLTERHDSR 218
OY 313 LHPDFVMAARIPENDODNDKVFYEFSETPSPDGSNHTVSNGVYCNADGGGRV 372
DB 219 WLNPKVVKVFWLPESENPDDDKYFEFFRETAVEAPALGRISVSRVQICNDVGQGRS 278
OY 373 LVNKRSTFLKARLVSVPGGAETHFDLEPFLMRKAGSLVVALFTSVAVPFGF 432
DB 279 LVNKRSTFLKARLVSVPGGEG-DTHFDLDVFLSSRDHRTPLLYAVFTSSIRGGS 337
OY 433 ANCVYHMDIWEVFNPAHNDGPOHNGPYGKVPFRPGVCSKMTAQDGRFGSTKD 492
DB 338 ANCVYSMDVRAFLGPRFAHNGEPHOMVSYGGRVPRPGMCSKTFG-----TFSTKD 393
OY 493 YPDEVLOFAHAPLTFMFPVRPRHGRPVLYKTHLAQQLHOIWDVREAEDEGTYYVIFLGD 352
DB 394 PPDVIOFARHAPLTFMFPVPRHGRPVLYKTHLAQQLHOIWDVREAEDEGTYYVIFLGD 352
OY 553 SGVLYKATLQAGSAPREVEVLEELQVFPVPTIMEISVKRMLYGSRLGVAQRL 612
DB 454 VGVLYKATLQAGSAPREVEVLEELQVFPVPTIMEISVKRMLYGSRLGVAQRL 612
OY 613 HOGETGTACAECCCLARDPCANDGCTHYRSLGKRRFRFODIRHNPALOCIGOSOE 672
DB 514 HRCAGHGRVCTECCCLARDPCANDGCTHYRSLGKRRFRFODIRHNPALOCIGOSOE 672
OY 673 EEAVALYATMYGTEHNSFLECPKSPQAAVRLRLDDEGPDQYKDERVLTHERG 732
DB 573 P-----ALHEHKEVGEVGSASFLECEPRSLQARVEMTFORAGVTAHTOVLEERTFARG 628
OY 733 LLEFRLSPFADATCTLEHGFSGOYVRLALVIVASOLDNF-PEPPPEPPARGGL 791
DB 629 LLEFRLSPFADATCTLEHGFSGOYVRLALVIVASOLDNF-PEPPPEPPARGGL 791
OY 792 ASTPKAMYNKIOLI-----GFANLPRVDECEVWCRCSTTECSGC-FRSGRCQAR 844
DB 685 -----PKLMTDFDLQVLEPPGGGGSANSLRM-----CRRPALQSLPSLRKGRNNR 731

RESULT 14

AA27127
ID AA27127 standard; Protein; 777 AA.

AA27127;

14-SEP-1999 (first entry)

Human brain tissue-derived polypeptide (clone OM007).

Brain tissue; human; bone marrow; umbilical cord venous endothelial cell;

recombinant; diagnosis; treatment.

Homo sapiens.

Key

Peptide

Protein

W09933873-A1.

08-JUL-1999.

25-DEC-1998; 98WO-JP05952.

26-DEC-1997; 97JP-0358811.

(ONOX) ONO PHARM CO LTD.

Fukushima D, Shibayama S, Tada H;

WPI; 1999-419088/35.

N-PSDB; AAX89112, AAX89113.

New adult human brain tissue-produced polypeptides useful for

diagnosis and treatment

Claim 1; Page 36-39; 86pp; Japanese.

The invention provides polypeptides (AA27127-Y27133) produced by human

adult brain tissue, human bone marrow or a human umbilical cord venous

endothelial cell. Host cells transformed with vectors comprising the

nucleic acids encoding the polypeptides are used for the recombinant

expression of the polypeptides. The polypeptides can be used in

diagnosis, treatment and basic studies, with wide applications in

treatment depending on the activity to be aimed at. Sequences

AAX89112-125 represent nucleic acids encoding the polypeptides.

Sequence 777 AA;

PR	10-SEP-1998	98US-0039806
PR	10-SEP-1998	98US-0039812
PR	10-SEP-1998	98US-0039815
PR	10-SEP-1998	98US-0039816
PR	15-SEP-1998	98US-0100385
PR	15-SEP-1998	98US-0100388
PR	15-SEP-1998	98US-0100390
PR	16-SEP-1998	98US-0100394
PR	16-SEP-1998	98US-0100627
PR	16-SEP-1998	98US-0100661
PR	16-SEP-1998	98US-0100662
PR	16-SEP-1998	98US-0100663
PR	16-SEP-1998	98US-0100684
PR	16-SEP-1998	98US-0100849
PR	18-SEP-1998	98US-0101014
PR	18-SEP-1998	98US-0101068
PR	18-SEP-1998	98US-0101071
PR	22-SEP-1998	98US-0101279
PR	23-SEP-1998	98US-0101471
PR	23-SEP-1998	98US-0101472
PR	23-SEP-1998	98US-0101474
PR	23-SEP-1998	98US-0101475
PR	23-SEP-1998	98US-0101476
PR	23-SEP-1998	98US-0101477
PR	23-SEP-1998	98US-0101479
PR	23-SEP-1998	98US-0101738
PR	24-SEP-1998	98US-0101743
PR	24-SEP-1998	98US-0101745
PR	24-SEP-1998	98US-0101915
PR	24-SEP-1998	98US-0101916
PR	29-SEP-1998	98US-0102207
PR	29-SEP-1998	98US-0102240
PR	29-SEP-1998	98US-0102307
PR	29-SEP-1998	98US-0102330
PR	29-SEP-1998	98US-0102331
PR	30-SEP-1998	98US-0102487
PR	30-SEP-1998	98US-0102570
PR	30-SEP-1998	98US-0102571
PR	30-SEP-1998	98US-0102684
PR	01-OCT-1998	98US-0102687
PR	01-OCT-1998	98US-0102965
PR	06-OCT-1998	98US-0103258
PR	06-OCT-1998	98US-0103449
PR	07-OCT-1998	98US-0103314
PR	07-OCT-1998	98US-0103318
PR	07-OCT-1998	98US-0103328
PR	07-OCT-1998	98US-0103395
PR	08-OCT-1998	98US-0103679
PR	08-OCT-1998	98US-0103671
PR	14-OCT-1998	98US-0104287
PR	20-OCT-1998	98US-0105000
PR	20-OCT-1998	98US-0105002
PR	21-OCT-1998	98US-0105104
PR	22-OCT-1998	98US-0105169
PR	22-OCT-1998	98US-0105266
PR	26-OCT-1998	98US-0105693
PR	26-OCT-1998	98US-0105694
PR	27-OCT-1998	98US-0105607
PR	27-OCT-1998	98US-0105681
PR	27-OCT-1998	98US-0105882
PR	27-OCT-1998	98US-0106062

```

PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106566.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106903.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 10-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108825.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
PA (GETH ) GENENTECH INC.
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX WPI: 2000-237871/20.
XX N-PSDB: AAA37109.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 176; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AA99340 to AA99462. Th
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 777 AA;
XX

```

[illegible]

```

Db      161 GYKKDITLFIKDOTNHLSEGRKLCFEDQOPASVYMDDELVLSTASDFLGKDTAFTRSLC
QY      301 P---RPAIRSD-SQSLLDHPRFVEMARIEPENSODNDKUYEFSEFSEVPSBDOGSNNHYV
Db      221 PTHDHHYIRTDISEHYMLNGAKFICGTFFIDYINPPDDKTYFFFRS--SOEBSISDKTI
QY      357 -SRVGRVCVNDAGORVULNKKMSTFLKALVCSYPCGBGAETHFDLEVLFLMPKAGKS
Db      279 LSRGRVCXKNDVGSQRSLINKMTTFKLARLCSIPSSDGDATYFDDELQIYLLTFPRDRN
QY      416 LEVYALFSTVASVEQGFAYCVYHMADIWEYFNENPFAHRODPOHOMVRYGKGKPFPRGVC
Db      339 PUYVGVFTTSSIFKSGSAYCVYSMADIRAVFNQRYAHKKSADHRWQYQIGRIPTPRGIC
QY      476 PSKTAOPRGPRFPGSTKQYPRVYLOFARHPLNWPVPRPHGRVPLVKTHLAOLHQIYWD
Db      399 PSK-TYPR-LISTRDPEDVATSEIKRRSHVYKSYVPAAGSPTEKRINVDRLQIYWD
QY      536 RVEADGTSTYVIFLGDSGSLVKVIALQAGSAEP---EEVLELOVFKVPTPTHEI
Db      456 HVAEAGQYUYVMFLGTGIDIGTVLKVASI-----SKEMNMEEVLELOLFKHSISLLNLEL
QY      593 SYKROILUYGSRIGVAVOLLIHQCEYTGTCACCCILARDPYCAMDASCTHYRPSLGGRRF
Db      512 SLKQOOLYIGSRDGLVOLTLHRCIDYUGKACACCILARDPYCAMDNACSRAPL-SKRA
QY      653 RRODIRHGNPALOC-----LGSOEEFAVGLVATVYGGFHNSTFPECILPKSPOAAVR
Db      571 RRODYKXGPRITOCMDIEDSISHEPDADEV-----IFGIEFNSTFPECILPKSQOATIK
QY      707 WILQPRGSDGPRQVUTDEVLVLTTEGGLFRRLSRFDAGTYCTTLEHGSQTVRLAV
Db      624 WTIQRSGDHRBELPDEITLITEGGLIRSLQKDSGMYCKAQOENHFIITYKTLPLV
QY      767 IYASQLDLPEPRKPEERPARGLASTPRPKAYKDILQILGFANLPRVDEYCYRVCYRG
Db      684 IENEMEN--TORAHEHGEQYVDLAES-PLRKDYIQLISSNF--SLDYCEQM---
QY      827 TTEGCGCFRSRRCQARGKAGL-ELGKMKSVYAHENKRPREV
Db      736 -----HREKROKNKGGPRVKKHNOEQEKKKRRNRHNRHDDLEIPRAV

```

Search completed: October 9, 2002, 14:21:52
Job time : 64 secs

Thu Oct 10 09:35:04 2002

us-09-813-290-2.1.rge

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
October 9, 2002, 19:18:09 ; Search time 2790 seconds
(without alignments)
6562.962 Million cell updates/sec

SUMMARIES

prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

28: em_vt:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Title: US-09-813-290-2
Sequence: 1 MACALAGKVPKMGSMVPMH.....KKMSRYVHNNPTREVEAT 875
Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delep 6.0, Delext 7.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_p2n.model -DEV=tbl
-Q=cnq2.1/USPTO.spool/US09813290/runat.09102002.094645.19122/app.query.fastq.1.1031
-DB=genEmbl -OPT=fastap -SUFFIX=rig -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr13290.ecgn.1.1828-efunat.09102002.094645.19122 -NCPU=6 -ICPU=3
-USER=US09813290.ecgn.1.1828-efunat.09102002.094645.19122 -WAIT -LONGLOG -DEV.TIMEOUT=120
-NO.XLPHY -NO.LMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*

Result No.	Score	% Match	Length	DB	ID	Description
1	4746	100.0	2628	6	AX253545	AX253545 Sequence
2	4746	100.0	3568	6	AX253549	AX253549 Sequence
3	4215	88.8	2349	9	AX253547	AX253547 Sequence
4	4215	88.8	4700	6	AB029496	AB029496 Homo sapi
5	4064.5	85.3	2340	6	AX155170	AX155170 Sequence
6	3913	82.4	4469	2	AK024425	AK024425 Homo sapi
7	3122.5	65.8	214046	2	AC0944021	AC0944021 Homo sapi
8	3110.5	65.5	123943	9	AC006208	AC006208 Homo sapi
9	3102.5	65.4	123943	6	AX155172	AX155172 Sequence
10	2282.5	48.1	229706	2	AC095672	AC095672 Rattus no
11	2056	43.3	2298	10	AF034744	AF034744 Mus muscu
12	2043.5	42.8	2615	5	AF022947	AF022947 Gallus ga
13	2033.5	42.5	6474	10	AB002329	AB002329 Human MRN
14	1955	41.2	3988	10	MM293947	MM293947 M. musculus
15	1955	41.2	4466	10	MM293947	MM293947 M. musculus
16	1904	40.1	5952	5	GG002528	GG002528 Gallus gall
17	1846.5	38.9	2319	10	MUSCSE	MUSCSE Mus musculus
18	1846	38.8	2319	10	MUSCSE	MUSCSE Mus musculus
19	1843	38.8	2319	10	MUSCSE	MUSCSE Mus musculus
20	1840	38.7	2913	5	AF083382	AF083382 Danio rer
21	1835.5	38.7	2709	6	AX207154	AX207154 Sequence
22	1835.5	38.6	2709	6	AX207154	AX207154 Sequence
23	1834	38.6	2709	6	AX207154	AX207154 Sequence
24	1830	38.6	2709	6	AX207154	AX207154 Sequence
25	1830	38.6	2709	6	AX207154	AX207154 Sequence
26	1830	38.6	2709	6	AX207154	AX207154 Sequence
27	1830	38.6	2709	6	AX207154	AX207154 Sequence
28	1797	37.9	2715	5	AF086761	AF086761 Homo sapien
29	1779	37.5	2919	5	AF124485	AF124485 Homo sapi
30	1773	37.3	2825	9	BC013975	BC013975 Homo sapi
31	1769	37.0	2825	9	BC013975	BC013975 Homo sapi
32	1755	36.5	2813	6	E27343	E27343 Novel poly
33	1733.5	36.3	3680	10	MUSSEMA11	MUSSEMA11
34	1721	35.8	1998	10	PC010976	PC010976 Human semap
35	1698.5	35.3	2499	9	HSU38276	HSU38276 Human clone
36	1673	35.1	3552	9	HSU38276	HSU38276 Human clone
37	1667.5	35.1	2872	10	HSU38276	HSU38276 Human clone
38	1667.5	35.1	2872	10	HSU38276	HSU38276 Human clone
39	1664.5	35.0	2265	10	AF080091	AF080091 Mus muscu
40	1659.5	34.7	2765	9	BC009113	BC009113 Homo sapi
41	1644.5	34.1	2765	9	BC009113	BC009113 Homo sapi
42	1644.5	34.1	2765	9	BC009113	BC009113 Homo sapi
43	1618.5	33.0	2336	5	AF022946	AF022946 Gallus ga
44	1568	32.9	2336	5	AF022946	AF022946 Gallus ga
45	1561.5	32.9	2336	5	AF022946	AF022946 Gallus ga

ALIGNMENTS

RESULT 1
AX253545
LOCUS AX253545
DEFINITION Sequence 1 from Patent WO0170806.
ACCESSION AX253545
VERSION AX253545.1 GI:16074048
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoideae; Homo.
REFERENCE
1 (bases 1 to 2628)
AUTHORS
Walke,D.N., Wilton, D.

D	b	2475	ACCCACTACGCGCCCCAGCCTTGGCAAGGCCGCCGTTCGCGCGGACGAGCATCCGCGCACGGC	2534
OY		661	AsnProAlaLeuGlnCysLeuGlnVglSerGlnIleuGluValGlyLeuValAla	680
D	b	2535	AACCCCTGCCCTCCATGCTGCCTGGCCAAAGCAGGAAGAAGAGGACAGTGGA	2594
OY		681	AlaThrMetValIyTryGlyTrpGlnHisSerSerThrPheLeuGlnCysLeuProLysSer	700
D	b	2595	GCCACCATGCTTACGCGACGAGGACCATATACCCTTCCCTGAGAGTCCGCCCACGCT	2654
OY		701	ProGlnAlaValAlaArgTrpLeuLeuArgProGlyLysArgLysGlyProAspGlnVal	720
D	b	2655	CCCCAGGCTGCTGCTGCCTGCTCTTTCCGAGAGGACGAGGATGAGGGGCTTACCAAGTG	2714
OY		721	LysThrAspGluArgValLeuHisThrGlnIleValGlyLeuLeuPheArgValLeuSerArg	740
D	b	2715	AGAGGAGACGAGCAGCTGTGCACACGAGGACGGGGCTGCTCTTCCGAGGCTTAAGCCT	2774
OY		741	PheAspAlaGlyThrTrpTrpCysThrThrLeuGlnHisGlyPheSerGlnTrpValAla	760
D	b	2775	TTCGATCCGGGCGACCTTACACCTGCACACCACTCTGGAGCATGGCTTCTCCACAGCTCGCT	2834
OY		761	ArgLeuAlaLeuValValIleValAlaSerGlnLeuAspAsnLeuPheProProLysPro	780
D	b	2835	CGGCTGGCTCTGCTGCTGATTTGTGGCTTACACCTGTGCACACCTTCCCTCCGAGCA	2894
OY		781	LysProGluGluProProAlaArgGlyGlyLeuAlaSerThrProProLysAlaTrpTyr	800
D	b	2895	AAGCCAGAGGAGCCGCCAACGCCGGGAGGCGCTGGCTTCCACCCACCAAGCCCTGTAC	2954
OY		801	LysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuProArgValAspGlyTrpCysGlu	820
D	b	2955	AAGGACATCTCTGACCTCATTTGGCTTCCCAACCTGCCCGGGTGATGATGACTGTAG	3014
OY		821	ArgValTrpCysArgIyLThrThrGlnCysSerGlyCysPheArgSerArgSerArgIy	840
D	b	3015	CCGCTGTGCTGACGGGACACCAAGMATGCTACAGCTGCTTCCGAGCGGAGCCGGGGG	3074
OY		841	LysGlnAlaArgIyLysSerTrpAlaGlyLeuGlnLeuGlyLysMetLysSerArg	860
D	b	3075	AAGCAGGCCAGGGGCAAGAGCTGGGACAGGCTGAGCAAGAGATGAAGAGACCGG	3134
OY		861	ValHisAlaGlnHisAsnArgThrProArgIleValGlnAlaThr	875
D	b	3135	GTCATGCCGAGCACAATCGACGCCCGGAGAGTGAGGCCACG	3179
RESULT 3				
AX253547				
LOCUS				
DEFINITION				
ACCESION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Alignment Scores:				
Fed. No.:				
Score:				

10

—

Alignment Scores:	
Pred. No.:	1,796-160
Score:	4215.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	88.81%
DB:	9
	Gaps:
	0
	Indels:
	0
	Mismatches:
	0
	Conservative:
	762
	Matches:
	4700
	Length:

ABU29496 (1-4700),

1	MetLarProSeValATrPrAlaLeuCYStrPrLeuLeuGly1yLeuLeuLeuHisG1yGly	113
Db	1 ATGGCCCCCTGGCGCTGGGCCCATTTGCTGGCTGCTAGGGGGCCCTCTGCTCATGGGGGT	113
QY	114 SerSeGlyProSePrGlo1yProSeValProAtGlyuAtGlyeSerTyuAtGlyaspLeu	60
Db	61 AGCTCTGGCCCCAGGCCCGGGCCCCAGTGTGGCCCGGCTGGGCTCTCTACAGAACCTC	133
QY	134 LeuSerAlaAsnArgSerAlaAlpheLeuGlyProGlnGlySerLeuHisnLeuGlnAla	120
Db	121 CTGTGTGGCCACCGCTGTGCCATCTTTCTGGGGCCCCAGGGCTCCCTGAACCTCCAGGCC	153
QY	154 MetTyLeuAsnArgTyuTyuArgaspArgLeuPheLeuGlyG1yLeuAspAlaLeuTySer	180
Db	181 ATGTACCTAGATGAGTATACCAAGACCGCGCTCTTTCTGGTGGGCTGGACCCCTCTACACT	173
QY	174 LeuArgLeuAsnArgAlaTrpProAspProArgGlyValLeuTrpProGlnProGly	240
Db	241 CTGGCGCTGGACCAAGCATGCCAGATCCCGGGAGAGTCTGTGGCCACCCAGCAGCAAGA	193
QY	194 GlnArgGluGluCyValArgTyuGlyArgaspProLeuTrpGlyCyValAsnArgPheVal	300
Db	301 CAGAGGAGAGAGTGTCTGCAGAAAGGAAGATCCTTGACAGAGTGGCCAACTTGCTG	213
QY	214 ArgValLeuGlnProHisAsnArgTrpHisLeuLeuAlaCysGlyTrpTrpAlaPheGln	360
Db	361 CGGGTCTACACCTCTCACACCGGACCCACTGCTACCCCTGTGGCACTGGGGCTTCCAG	233
QY	234 ProTrpCysAlaLeuLeuTrpValG1yHisArgGlyGlnHisValLeuHisLeuGlyPro	420
Db	421 CCCACCTGTGGCTTCATCAGATTGGCCACCGTGGAGGAGCATGTGCTCCACCGAGGCT	253
QY	254 GlySerValGlySerGlyArgGlyArgCysProHisG1yProSeArgTrpPheAlaSer	480
Db	481 GGCAGTGTGGAAAGTGGCCGGGGCGGCTGGCTCCACAGACCCACCGCTCCCTTGTCCAGC	273
QY	274 ThrPheLeuAspGlyGluLeuTyuTrpGlyLeuTrpAlaAspPheLeuGlyArgGlyAla	540
Db	541 ACCTTCATTAACGGGGAGCTGTACAGGGTCTCATCTGTGACATTCCTCTGGCGAGAGGCC	293
QY	294 MetLeuPheArgSerGlyGlyProArgProAlaLeuArgSerAspSerAspGlnSerLeu	600
Db	601 ATGATCTTCGAAAGTGGAGGCTCTGGGGCAGCTGTGCTGTCCACTGTGACCAAGAGTCTC	313
QY	314 LeuHisAspProArgPheValMetAlaAlaTrpLeuProGlyuAsnSerAspGlnAspAsn	660
Db	661 TTTCACAGACCCCCGGTGTGTGATGGCCCCCGGATCCCTAGAACCTTCAACCGAGCAAT	333
QY	334 AspLysValTyuPhePhePheSerGlyuTrpValProSerProAspGlyGlySerHis	720
Db	721 GACCAAGGTGTACTTCTTCTCGGAAGACGTCCTCCGCGCGCATGTGGTCTGAAACCAT	353
QY	354 ValTrpValSerArgValGlyArgValCysValAsnAspAlaGlyGlyGlnArgValLeu	780
Db	781 GTCACTGTCAACCGCGGGGCGCGCTGTGGTGAATGATGACTGGGGGCCAGCGGTGTCTG	373
QY	374 ValAsnLysTrpSerTrpPheLeuLysValArgLeuValCysSerValProGlyProGly	840
Db	841 GTGACAAAGAGGCACTTTCTCCAGAGCCCAAGCGTGTCTGTGGTCCCGGCTGGT	393
QY	394 G1yAlaG1yTrpHisPheAspGlnLeuGlnAspValPheLeuLeuTrpProLysAlaGly	900
		413

Db 501 GGTCCGAGACCACTTTGACCAAGTAGAGATGTTCTCTGTCGCCCAAGCCGG 960
 Oy 414 LysSerLeuGluValTyrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAla 433
 Oy 961 AAGACCTCCAGAGGTATGAGCCCTGTCAGCAACCTCAGTACGGCTGTTCAGGGCTTCGCC 1020
 Oy 434 ValGlyValTyrHisMetAlaAspIleTrrGluValPheAsnGlyProPheAlaHisArg 453
 Db 1021 GTCCTGTGTACCATGAGACATCTGGAGAGTTCACGGGGCCCTTGGCCACCGCA 1080
 Oy 454 AspGlyProGlnHisGlnTrrGlyProTyrGlyLysValProPheProAspProGly 473
 Db 1081 GATGGGCTCAGACACAGTGGGGGCTTATGGGGCAAGGTGCTTCCTCCCTCCGCTGGC 1140
 Oy 474 ValGlyProSerLeuMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTyr 493
 Db 1141 GGTGCCCCAGAACATGACCCGACAGCGAGCGGCTTTGGACACCAAGAGACTAC 1200
 Oy 494 ProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrrProValArgPro 513
 Db 1201 CCAGATGAGAGTCTGCAATTTGCCAGCCACCCCTCATGTTGTGGCTGTGGCGCTT 1260
 Oy 514 ATGHSGLAArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleVal 533
 Db 1261 CGCATGAGCCGCTCTCTCTTCACAGACCCACTGGCCAGAGCTACACAGATCGTG 1320
 Oy 534 ValAspArgValGluAlaGluAspGlyThrTyrAspValIlePheLeuGlyThrAspSer 553
 Db 1321 GTGAGCCGGTGGAGCAGAGAGTGGACCTACAGATGTCTATTTCCGGGAGCTACCA 1380
 Oy 554 GlySerValLeuLysValIleAlaLeuGlnAlaGlyLysSerAlaGluProGluGluVal 573
 Db 1381 GGGTGTGTGTCAAAGTCACTGCTCTCCAGGAGGGGCTCAGCTGAACCTGAGGAAGTG 1440
 Oy 574 ValLeuGluGluLeuGlnValPheLysValProThrProIleThrGluMetGluLeuSer 593
 Db 1441 GTTCTGGAGAGCTCCAGGTGTAAAGTGCACACACTATCACGGAATGAGACTTCT 1500
 Oy 594 ValLysArgGlnMetLeuTyrValGlySerArgLeuGlyValAlaGlnLeuArgLeuHis 613
 Db 1501 GTCAAAAGGCAATGCTATAGTGGCTCTGGCTGGGTGGTGGCCAGCTGGCTGGCAC 1560
 Oy 614 GlnGluGluThrTyrGlyThrAlaCysAlaGluCysCysLeuAlaArgAspProTyrCys 633
 Db 1561 CAATGTAGACTTACGGCAGCTGCTGTGCAAGTGTCTGCGCCGGAGCCATATCTGT 1620
 Oy 634 AlaTrrAspGlyAlaSerCysThrHisTyrArgProSerLeuGlyLysArgArgPheArg 653
 Db 1621 GCTGGGATGGTCTCTCTTACCCACTACCGCCCTGGCAAGCGGGGTTCGCG 1680
 Oy 654 ArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCysLeuGlyGlnSerGlnGlu 673
 Db 1681 CGGAGGACATCCGGCAGCGGCAACCTGCTGCGAGTGTGGCCAGAGCCAGGAAGA 1740
 Oy 674 GluAlaValGlyLeuValAlaAlaThrMetValTyrGlyThrGluHisAsnSerThrPhe 693
 Db 1741 GAGCAGAGGGAGCTTGTGCGACGCCATGTGCTACGCGCAGGAGCAACATGACCTTC 1800
 Oy 694 LeuGluCysLeuProLysSerProGlnAlaValArgTrrPheLeuGlnArgProGly 713
 Db 1801 CTGGAGTCTCCGCCCAAGTCTCCAGGCTGCTGGCTGGCTCTTGGCAGAGGCCAGGG 1860
 Oy 714 AspGluGlyProAspGluValLysThrAspGluArgValLeuHisThrGluArgGlyLeu 733
 Db 1861 GATGAGGGGCTGACAGTGAAGACGGAGCGAGAGTCTTGACACAGGAGCGGGGCTG 1920
 Oy 734 LeuPheArgArgLeuSerArgPheAspAlaGlyThrTyrThrCysThrTrpIleGluHis 763
 Db 1921 CTGTTCCGAGGCTTACCGGTTTCATGGCGGACCTACACCTGCAACCTCTGGAGCAT 1980
 Oy 754 GlyPheSerGlnThrValValArgLeuAlaLeuValValIleValAlaSerGlnLeuAsp 773
 Db 1981 GGCTTCTCCACAGTGTGTCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040

Oy 774 AsnLeuPheProGluProLysProGluGluProAlaArgGlyGlyLeuAlaSer 793
 Db 2041 AACCTGTCTCTCCGAGACCAAGCCAGAGAGCCCCAGCCGGGAGGCTGTGCTTCC 2100
 Oy 794 ThrProPolysAlaTrrPrrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuPro 813
 Db 2101 ACCCCACCAAGGCTGTGTACAGGACATCTGCTGCTCATGTGCTTGGCCACCTGCCC 2160
 Oy 814 ArgValAspGluTyrCysGluArgValTrrPrrCysArgGlyThrThrGluCysSerGlyCys 833
 Db 2161 CGGGTGTGTGAGTCTGTACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
 Oy 834 PheArgSerArgSerArgGlyLysGlnAlaArgGlyLysSerTrrPrrAlaGlyLeuGluLeu 853
 Db 2221 TTCCGAGACCCGAGCGCGGGGAGCAGAGGCCAGGCAAGACTGGGGCAGGCTGGAGCTA 2280
 Oy 854 GlyLysLysMetLysSerArgValHisAlaGlnHisAsnArgThrProArgGluValGlu 873
 Db 2281 GGCAAGAAGATGMAAGCGCGGTGATGCCGAGCACATTCGAGCGCCCGGAGGTGGAG 2340
 Oy 874 AlaThr 875
 Db 2341 GCCACG 2346
 RESULT 5
 AX155170 2340 bp DNA linear PAT 22-JUN-2001
 LOCUS Sequence 1 from Patent WO0140278.
 DEFINITION AX155170
 ACCESSION AX155170 GI:14536655
 VERSION AX155170.1 GI:14536655
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2340)
 TITLE Human semaphorin zsmf-16
 AUTHORS Holloway, J.L. and Foley, K.P.
 JOURNAL Human semaphorin zsmf-16
 JOURNAL Patent: WO 0140278-A 1 07-JUN-2001;
 ZymoGenetics, Inc. (US)
 FEATURES
 source 1..2340
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2340
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC42673.1"
 /db_xref="GI:14536656"
 /translation="MAPSNAWICMLIGILLHGGSSGSPSPSVRLRLSYRGAMVRK
 PSTMMETFSRYLSSNRSAITPLGDSGLNQAMTYDEYRDLFGGLDALYSLRD
 QAMPDPRRETCANFVRVLOPHNRTNHLACGTGAOPLCALITYGHRBHAIIEPGSV
 ESGRGCPHEPSRPFSTFIDGELYTLTDFLREAMITFRSGRPALRSDQSLL
 HDPRFMAAI PENSDQNDKUYEFFEETVPSDPGSDNHYVSVRGVANDAGGRV
 LVNKSFTFLKARLYCSVPGAGETNHDQLEDFVLLPKKSKSLVLAFTSVAVFG
 GFAYCVYHMDIIVLEVENGFPAHRDGPQHOHPYCGKGFPPGVCPSKMTAQPRPG
 STDYIDPEVLPQRAHRLMFWPVRBRGRPLVTLHAQOLHOLIVDFVEAEDGYDY
 IFLGDSGSVLKVIATDAGSSABREYVLEIYQFKYPTTTEISVKKRMLYVGR
 LGVAOLRLHOCETVGTACARCCUARDYCAMDASCTHYRSLKRRPRRQDIRGNP
 ALQCLGOSDEEAGLVAAATVGTENSTFLBCLPSPQAAVRLRORPDEGDOY
 KTDERVLHTERGLEFRSLSFDAVDTCTLLEHSPVSRVLAIVTVAISDLNLEPP
 EKPDEEPPARGGLASTPRAKYADLTLLIGFANLPRADEVYCEYRWCRCGTBCSGCFS
 RSKCKORGRSMGLTEIGKKMSVHAEHNTPREVENT"
 BASE COUNT 406 a 748 c 743 g 443 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,18e-154 Length: 2340
 Score: 4046.50 Matches: 760
 Percent Similarity: 95.00% Conservative: 0
 Best Local Similarity: 95.00% Mismatches: 1
 Query Match: 85.26% Indels: 39

DB: 6 Gaps: 2

US-09-813-290-2 (1-875) x AX155170 (1-2340)

QY 94 MetAlaProSerAlaIleTrrAlaIleCysTrrPLeuLeuGlyLeuLeuLeuHisGlyGly 113

1 ATGGCCCCCTCGGGCCGATTTCTGCTCTAGGGGGCCCTCTGCTCATGGGGGT 60

QY 114 SerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTyrArgAsp 132

61 AGCTGTGGCCCCAGCCCGCCCAATGTGCCCGCTCGGGCTCTCTACCGAGAACCC 120

QY 133 -----LeuLeuSer 135

121 ATGTCCGAAAGCCCTCCAGCACCATGTGATGAAACATTTTCCAGATACCTCTGTCT 180

QY 136 AlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTyr 155

181 GCCAAGCCCTCTGCTCATCTTTTGGGGCCCCAGGGCTCCCTGAACCTCCAGGCCATGTAC 240

QY 156 LeuAspGlyTyrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArg 175

241 CTAGATGATGATCCAGAGACCGCTCTTTCTGGGTGGCTGGACCCCTCTACTCTCTGGG 300

QY 176 LeuAspGlnAlaTrrProAspProArgGlyValLeuTrrProGlnProGlyGlnArg 195

301 CTGGACCAAGGATGGCCAGATCCCGGGAG----- 330

QY 196 GlnGluCysValArgGlyGlyArgAspProLeuThrGluCysAlaAsnPheValArgVal 215

331 -----ACAGAGTGGCCCACTTCTGTGGGGTG 357

QY 216 LeuGlnProHisAsnArgTrrHisLeuLeuAlaCysGlyTrrGlyAlaPheGlnProThr 235

358 CTACAGCCTCAACAACCGGACCCACTGCTAGCCTGTGGCAGTGGGGCTTCACAGCCACC 417

QY 236 CysAlaLeuLeuThrValGlyHisArgGlyGlnHisValLeuHisLeuGlnProGlySer 255

418 TGTGCCCTCATACAGTGTGGCCAGCGTGGAGCATGTCTCTCAGCTGGAGCCTGGAGCT 477

QY 256 ValGlnSerGlyArgGlyArgCysProHisGlnProSerArgProPheAlaSerThrPhe 275

478 GTGGAAAGTGGCCGGGGGGGCTGCTCTACAGACCCAGCCGCTCTTCCAGCACCTTC 537

QY 276 IleAspGlyLeuLeuTrrGlyLeuThrAlaAspPheLeuGlyArgGlnAlaMetTyr 295

538 ATGACGGGAGGCTGACAGCGGTCTCACTCTGCTCTGGGGGAGAGGCCATGATCT 597

QY 296 PheArgSerGlyGlyProArgProAlaLeuArgSerAspSerAspGlnSerLeuLeuHis 315

598 TTCCGAAGTGGAGGCTCTCGGCGCTCTGCTGCTCCGACTCGACACAGAGTCTCTGCAC 657

QY 316 AspProArgPheValMetAlaAlaArgIleProGlnAsnSerAspGlnAspAsnAspGly 335

658 GACCCCGGTTTGTGATGGCCCGCCGATCCCTGAGACTCTGACAGACAGACAAATGACAAG 717

QY 336 ValTyrPhePhePheSerGlnTrrValProSerProAspGlyGlySerAsnHisValThr 355

718 GTCTACTTCTTCTCTCGAGACGCTCCCTCGCCCATGTGGTGGCTCAACCATGTCACT 777

QY 356 ValSerArgValGlyArgValCysValAsnAspAlaGlyGlyGlnArgValLeuValAsn 375

778 GTTCAGCCGGGTGGGGCGGCTGCTGTAATGATCTGCGGGCCAGCGGCTGTGTGAAC 837

QY 376 LysTrrSerThrPheLeuLysAlaArgLeuValCysSerValProGlyGlyAla 395

838 AAATGGAGCACTTCTCTCAAGGCCAGCTGTGTGCTGCTGCGCCCGCTGTGTGTGGC 897

QY 396 GlnTrrHisPheAspGlnLeuGlnAspValPheLeuLeuTrrProLysAlaGlyLysSer 415

898 GAGACCCCACTTTGACAGGTAGAGGATGTCTCTGCTGTGGCCCAAGCGGGAGAGGC 957

QY 416 LeuGlnValTrrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCys 435

DB 958 CTCGAGGTGTACGGCGCTTTCAGCACCGCTCAGTGGCGTGTTCAGAGGCTTCCCGCTCT 1017

QY 436 ValTrrHisMetAlaAspIleTrrGlyValPheAsnGlyProPheAlaHisArgAspGly 455

1018 GTCTACCACTGGCAGACATCTGGAGAGTTTTCACAGGGGCTTTGGCCACAGCAATGGG 1077

QY 456 ProGlnHisGlnTrrGlyProTrrGlyGlyLysValProPheProArgProGlyValCys 475

1078 CCTCAGACACAGTGGGGGCGCTTATGGGGGCAAGGTGGCTTCCCTCGCGCTGTGGCTTGC 1137

QY 476 ProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTrrProAsp 495

1138 CCGACGAAAGTGAACCGACAGCCAGAGGCGCTTTTGGACACACCAAGACATCCAGAT 1197

QY 496 GlnValLeuGlnPheAlaArgAlaHisProLeuMetPheTrrProValArgProArgHis 515

1198 GAGGTCTGCTGAGTTTGGCCGAGCCACCCCTCATGTCTTCTGGCTGTGGCTGTGACAT 1257

QY 516 GlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleValValAsp 535

1258 GGCCGCGCTGTCTCTTTCAGACCCACTGGCCAGCAGCTACACAGATGTGTGTGGAC 1317

QY 536 ArgValGlnAlaGlnAspGlyTrrTyrAspValIlePheLeuGlyTrrAspSerGlySer 555

1318 CGCGTGGAGGAGAGGATGGAGCTAGATGTCTTTTCTGGGGAGCTAGCTAGAGTCT 1377

QY 556 ValLeuLysValIleAlaLeuGlnAlaGlyGlySerAlaGlnProGlnGlnValValLeu 575

1378 GTGCTCAAGATATCTCTCTCCAGGACAGGGGGCTCAGCTGAACTGAGGAAGTGTGTCTG 1437

QY 576 GlnGluLeuGlnValPheLeuValProThrProIleThrGlnMetGlnIleSerValLys 595

1438 GAGGAGCTTCAGGTGTTTAAGTGTCCAAACCTTATCCSSAAATGGAGATCTGTGTAAA 1497

QY 596 ArgGlnMetLeuTrrValGlySerArgLeuGlyValAlaGlnLeuAspLeuHisGlnCys 615

1498 AGGCAAAATGCTTAGCTGGCTCTCGGCTGGGTGGCCAGCTGAGCGGGCTGCACAAATGT 1557

QY 616 GlnTrrTyrGlyThrAlaCysAlaGlnCysCysLeuAlaArgAspProTrrCysAlaTrr 635

1558 GAGACTTACGGCAGCTGCTGTGACAGATGTGCTGGCCGGGACCCATATCTGTGTGG 1617

QY 636 AspGlyAlaSerCysThrHisTrrArgProSerLeuGlyLysArgArgPheArgGln 655

1618 GATGTGCTCTCTTACCACTTACCGCCCAAGCTTGGCAAGCGCGGCTTCCGCGGACG 1677

QY 656 AspIleArgHisGlyAsnProAlaLeuGlnCysLeuGlyGlnSerGlnGlnGlnAla 675

1678 GACATCCGGCAGCGCAACCTTCCCTGCAGTGCTGGCCAGAGCCAGGAAGAGGCA 1737

QY 676 ValGlyLeuValAlaAlaThrMetValTrrGlyThrGlnHisAsnSerThrPheLeuGln 695

1738 GTGGAGACTTGTGGAGCAGCACTATGTGTGCGGAGCAGACACAAATGACACTTCTGTGAG 1797

QY 696 CysLeuProLysSerProGlnAlaAlaValArgTrrPheLeuGlnArgProGlyAspGly 715

1798 TGGCTGGCCCAAGTCTCCCAAGGCTGTGTGCTGTGCTGTGAGAGGCCAGGGATGAG 1857

QY 716 GlnProAspGlnValLysThrAspGlyArgValLeuHisThrGlnArgGlyLeuLeuPhe 735

1858 GGGCTGTACCAAGGTGAACAGCAGACGAGCTTGTACACAGAGAGGGGGGTGTGTTC 1917

QY 736 ArgArgLeuSerArgPheAspAlaGlyTrrTyrThrCysThrThrLeuGlnHisGlyPhe 755

1918 CGAGGCTTACCCCTTTCGATGGGGGCACTTACACTGTGACACACTGTGAGATGTGCTTC 1977

QY 756 SerGlnThrValAlaArgLeuAlaLeuValIleValAlaSerGlnLeuAsnLeu 775

1978 TCCCAAGATGTGTGGCTGT 2037

QY 776 PheProGlnProLysProGlnGlnProProAlaArgGlyGlyLeuLysSerThrPro 795

FEATURES		14709		58095: contig of 43387 bp in length
*		58096		58195: gap of unknown length
*		58196		125935: gap of 67740 bp in length
*		125936		126035: gap of unknown length
* 126036		214046: contig of 88011 bp in length.		
Location/Qualifiers		1. 214046		
source		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/chromosome="3"		
		/clone="RP11-801L18"		
misc_feature		1. 14608		
		/note="assembly_name:Contig24"		
misc_feature		14709. 58095		
		/note="assembly_name:Contig25"		
misc_feature		58196. 125935		
		/note="assembly_name:Contig26"		
misc_feature		126036. 214046		
		/note="assembly_name:Contig27"		
BASE COUNT		50192 a 57272 c 57247 g 48853 t		482 others
ORIGIN				
Alignment Scores:		3.09e-115		Length: 214046
Pred. No.:		3122.50		Matches: 866
Score:		26.93%		Conservative: 0
Percent Similarity:		26.93%		Mismatches: 9
Best local Similarity:		65.79%		Indels: 2350
Query Match:		2		Gaps: 15
DB:				
US-09-813-290-2 (1-875) x AC094021 (1-214046)				
QY 1		MetlAcysAlaAlaAlaGlyValPheProMetGlySerTrpProValTrpHisLys		20
Db 130817		ATGGCTGTGGCTTACGTGGGAAAGGTCTTCCAAATGGGAGCTGGCCAGTGTGGCACAAA		130876
QY 21		SerLeuHisTrpAlaAsnLysValGluGlyAlaAlaGlyValArgGlnGlyProSer		40
Db 130877		AGCCTGACCTGGCCCAACAGGTGAGAGAGACGCGAGTGTGAGCGCAAGGCCAGC		130936
QY 41		LeuLeuLeuSerSerAlaProLeuProAlaGlnAspTrpValGluProLeuProTyrLys		60
Db 130937		CTCTTCTCTCTCTCGGCCCTCTCTCCGCCAGACTGGGTGAGCAGCTTATTAAG		130996
QY 61		TrpTrpProGlyLysSerArgAlaAsnTyrAsnArgArgProAlaGlyProGluGly		80
Db 130997		TGTGTGGCTGTGTGGCAGACAACTAACCGCGCGCCAGCGGAGCCAGAGGGCGGC		131056
QY 81		SerAlaGlyArgArgGlnArgCysProGlnPheProSerMetAlaProSerAlaTrpAla		100
Db 131057		TCTCGAGCGAGCGGCGGAGCGGCTCAGTTCCACAGCATGGCCCCCTGGGCTGGGCC		131116
QY 101		IleCysTrpLeuLeuGlyLysLeuLeuHisGlyLysSerGlyProSerProGly		120
Db 131117		ATTGTCTGGCTGTGTAGGGGGCTCTCTGCTCCATGGGGTAGCTGTGGCCCCAGCCGGGC		131176
QY 121		ProSerValProArgLeuArgLeuSerTyrArg		131
Db 131177		CCCAAGTGGCCCCGCTGGGCTCTCTACCGAGGTACCAAGCCCTGCGCTGTTTG		131236
QY 131				131
Db 131237		TGTGATGTGGCCTTTTGTGTGTGTGGCTGTGGTGGTGTGATGTGCGGTGTTGTGT		131296
QY 131				131
Db 131297		GCATGTGCCCTGT		131356
QY 131				131
Db 131357		TGTGCGAGATTGGCTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		131416

QY 131	-----	-----	131
Db 131417	TGTTTGGGTCATGTGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		131476
QY 131	-----	-----	131
Db 131477	TTGTGTGCGTGTGCGGAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		131536
QY 131	-----	-----	131
Db 131537	TGTGCGTGTGCGCATGT		131596
QY 131	-----	-----	131
Db 131597	GCGTGTGCGCTGT		131656
QY 131	-----	-----	131
Db 131657	GCCTTCTCCGTGACACAGCTCTCGGGGAGTGGAGGAGGAGGACAGGCTCTGGAGACCA		131716
QY 131	-----	-----	131
Db 131717	GGGTGCGGTTGGAGGACACCTGAGGGGAGTGCAGAGAGGAGACAGATGAGACCTGA		131776
QY 131	-----	-----	131
Db 131777	GAGAGCTCAATAAAGATGGGAGAGACAGAGAGAGGAGGAGGAGGAGAGTGTGACAG		131836
QY 131	-----	-----	131
Db 131837	AGAGGTTGGGCTGAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		131896
QY 131	-----	-----	131
Db 131897	ACGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC		131956
QY 131	-----	-----	131
Db 131957	CCCTTGACGCCACTTCCAGAGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT		132016
QY 131	-----	-----	131
Db 132017	CCAGCCAGGGGAGTTTCACTTCCCTTTTGTGACATCCAGCCTGGGGTTCCTGTGT		132076
QY 131	-----	-----	131
Db 132077	TCTACCTAGACATGTGAGCGCGCTCAGGCTCCAGGATGGGCCAGGAGGCCGAG		132136
QY 131	-----	-----	131
Db 132137	ACAGCTTGTGAGATCTTCTGTGAGAAACAGAGGGGTGACACCTTCCCTGGGAA		132196
QY 131	-----	-----	131
Db 132197	GCCCACTTGTGTGGGAGATGCGCTGTCTCTTGAGAGAGCTGAGGGGAATGTGTCTT		132256
QY 131	-----	-----	131
Db 132257	GTTCCGGAGCTGT		132316
QY 131	-----	-----	131
Db 132317	AAGTCCCAAGTACTGTATCATCATGAGGCTTCCCGAAGAGACGTTGATGATTT		132376
QY 131	-----	-----	131
Db 132377	TGGGGTGAATGCGT		132436
QY 131	-----	-----	131
Db 132437	GCCTTCCCAAGTGGGAGCAGCAGCCTTCTTGCACATCTCTACTGCCCCATTACCTCC		132496
QY 131	-----	-----	131

```
b 132497 ACCTCCTAGCAAGCCCGGAGGACCTACCTAGAGATCTCAACTCTGTGGACCATGTC 132556
y 131 -----
b 132557 CCCAGACAGCTCCCGCTCTCANTCTGTGGGTGGGANTAGCAGCATCTGTGGGAGAAAA 132616
y 131 -----
b 132617 TTGTCTTTGGCCGCATCCCTGAGCTGAGAGTGGGAGGCCCTTCCCGCCATACCTTCCCTT 132676
y 131 -----
b 132677 TTCTGACCCAGGAGCCATGTGCCAGGGGTCCCTCTGTGGAAAGTTCGCTGTGGCC 132736
y 131 -----
b 132737 CACCCTCAGCTGAGCCAGGCGGCTGCAAAAGACTCCCGCTTCTCCAGAGATCTTCCA 132796
y 131 -----
b 132797 GCACCTTGTGCTAGTTTCTGTCTGTGGGCTTCTTCCCTGGGACTGGTAGGGACA 132856
y 131 -----
b 132857 GATGGAGAGTGGAGAAAGATGTAGCTGATGATGAAGGCTAAGAGATGAGAGGTGGGA 132916
y 131 -----
b 132917 GGTAGCCCCAGGAGTAGGCTGGGCCCAGGTGGAGGCTGTAGCCTCCGCTCCACCTGTC 132976
y 131 -----
b 132977 CCCAGAGCCGTCGTGCCAAGAGCCTTCCAGACCATGTGATGGAAACATTTCCAGATG 133036
y 131 -----
b 133037 TGAAGTGTCAAGGGGTGATGCTCAGCCGGGGCCTCCACTCTGTGGCTGGGGA 133096
y 132 -----
b 133097 CTAGGCTAGCTGTGGCCCTGATCTGCTCCCTCCCTACATCTCTCTAGACCTCTGTCTGC 133156
y 1331 -----
b 133157 CAACCGCTCTGCCATCTTCTGGGCCCCCAGGGCTCCCTGAACCTCCAGGGCATGTACCT 133216
y 136 aasnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetLysIle 156
b 133217 AGATAGTACCGAGACCGCTCTTCTGTGGGTGGCTGACGCCCTCTACTCTCTCGGCT 133276
y 176 uasPgluIyArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgIle 176
b 133277 GGACCAGGCAATGGCCAGATCCCGGGAGGT -GAGCTGAGACTGGCAGGGAAGAGGTCTCT 133335
y 186 -----
b 133336 TTGAGAACTGGGCACTCCAGATGGAGGCGCAGATGGGAAACTGAGCGCTGTGATCAG 133395
y 187 -----
b 133396 GGGAGGACCTGAGTCAGCGGGGCTGTGTCCAGGTCTCTGTGGCCACCGCAGCAGACA 133455
y 194 natrgluGluCysValArgIysGlyArgAspProLeu----- 206
b 133456 GAGGAGAGAGTGTGTTCGAAAAGAGAGATCTCTT -GTTAGTGTCTGTGGAGAGGATG 133514
y 206 -----
b 133515 GTCCCGACCTGACACCTGCTGACCCCGCAGACAATCCGTGACACGCCCCAGTA 133574
y 206 -----
b 133575 CCTCTACAGGGGCAATTGGAGAGATCTTGAGGGCTTTTGAAGTCTGTGGGAATGTC 133634
y 206 -----
b 133635 ACACCCACCCCTCCCGCAGAGCTGACCGCCAGACAGGGCTGAGCCTTGGCCCCAAGGA 133694
y 207 -----
b 133695 GATGGTGGGACCCCTCTCACCCTCAACCCACCCCTCTCTTCTCCCGCAGACAGAGTG 133754
y 209 -----
b 133755 salAsnPheValArgValLeuGlnProHisAsnArgThrHisLeuAlaCysGlyThr 229
b 133755 CCGCAACTTCGTGCGGGTGTCTACAGCCTCAACACCGGACCCACTGTAGCTGTGGCAC 133814
y 229 rGlyAlaPheGlnProThrCysAlaLeuIleThrValGlyHisArgGlyGlu----- 246
b 133815 TGGGGCTTTCAGACGACCATCGTGCCTCATCAGATTGGCCACCGTGGGGA -GTTGAGCC 133873
y 246 -----
b 133874 TGGGCCAGGCCCCACAAGGACTGGAGGCACTGACATTTCTCTCCCTGCTGACAT 133933
y 246 -----
b 133934 TCCCATTTAGCCTGGGTAGGAGACAGAGGGCGCTGAAGGCGCCTTGTGTGGG 133993
y 246 -----
b 133994 GAGCTGACCCCATTTAGAGGGGCGCTGATGAGAGCGCTTCGCCCCACCGCTCTCTG 134053
y 246 -----
b 134054 GCGGCAAGGAGGCGTCTGCTGGGTGGGTGGCTGTGCTGTCTCTCCAGGTGACCC 134113
y 247 -----
b 134114 CCTCCCTCCCTCTCTCCCTCCACTGACATGTGCTCCACTGGAGCCTGGAGCTGTGA 134173
y 257 userGlyArgGlyArgCysProHisGlyProSerArgProPheAlaSerThrPheIle-- 276
b 134174 AAGTGGCCGGGGGGGTGGCTCCTACAGAGCCAGCGTCCCTTGTGCAGACCTTCAT -AG 134232
y 276 -----
b 134233 GTGGGTGATGCCAGGCCAGGGGAGAGAAACAGGAGGGGTGCGAGCTATGTCACTGG 134292
y 276 -----
b 134293 GAAGGAGGAGCCCGGCTGCTGGGCTGGGCTCCCTCCCTGACTCTGTGCCCAACCCC 134352
y 277 -----
b 134353 CACCCAGAGCGGGAGCTGTACACGGGTCTCAGCTGCTGACTTCTCGGGGCGAGAGGCCAT 134412
y 294 tIlePheArgSerGlyGlyProArgProAlaLeuArgSerAspSerAspGlnSerLeu 314
b 134413 GATCTTCGAAAGTGAAGTCTGCTGGCCAGCTCTGCTCCAGCTGTGACAGAGTCTCTT 134472
y 314 uHisAsp----- 316
b 134473 GCACGG -TGAGGCTGCCCCCTGGAACTGGCAGGGGTGAAGAGGCTACCTTCCAGCTAGG 134531
y 316 -----
b 134532 ACACAGGCTCCAGTGGAGTCCCTGCCCCCATTTCTCTCCCACTGGGTCTACTCCA 134591
y 317 -----
b 134592 TCTTCCCTGTGGGGCCCTCCACCCAGCATGACCTGTACACCTTCCAGAGACCCCGGT 134651
y 319 -----
b 134652 TGTGATGGCGCGCGGATCCCTGAGAACTCTGACACGAGACAAATGACAGAGGTACTTCTT 134711
y 319 -----
```

OY	339	ePhSeRgluThrValProSerPrioAspGlyLySerAsnHisValThrValSerAlgya	359
Db	134712	CTTCTCGAGAGACGGTCCTCCCTGCCGATGGGTGCAGAACCATGACTCATTGACGCCGT	134771
OY	359	lGlAryrVal----- 	362
Db	134772	GGGCCGCTTGCGGTGAAGAGCTGGGAAGGGGTGTGAGTTCAAGTCCCTGCTACAC	134831
OY	362	-----	362
Db	134832	CCTCCATTCCCAGACCTTTCCTCCCCACCCTTGCCAGCAGTAGGAAACTGAGCCAA	134891
OY	362	-----	362
Db	134892	GGATGCAAAAGCTGACATGCATACCTCTCCCGAGTGGGTGAGGTTAGAAGAATGGATGCT	134951
OY	363	-----CysValAsnAspAl 	367
Db	134952	GGGACCCATCTCGTGGGGGCCCAAGGGTGACTCTGGCCCCCTTAATCTGCAG-ANTGATGC	135010
OY	367	aGlyGlylnarValLeuValAsnLysTrpSerTrhrPheLeuLysAlArglyLeuValCy	387
Db	135011	TGGGGCCAGGGGGGTGGTGTAAACAAAGGACACTTTCCTCAAAGGCCAGCGTGGTCTTG	135070
OY	387	sSerValProGlyProGlyGlyAlaGluThrHisPheAspGlnLeu-----	402
Db	135071	CTCGGTGCCCGGCCCTGGTGGTGGCCGAGACCACCTTTGACCACT-AGTGAAGGGCTTG	135129
OY	402	-----	402
Db	135130	CCAAGCAGGCTTAGGGCTCGGGAGTGGGGAGGGCTGAGACTGTGACTCAGGGCATTTGG	135189
OY	403	-----GluAspValPheLeuLeuTrpPr 	410
Db	135190	CGGTGGCCATGACACTCATCTCTTCTTCTACCCAGAGAGATGTGTTCTCTGCTGTGGCC	135249
OY	410	oLysAlaGlyLysSerLeuGluValTyraAlaLeuPheSerThrVal-----	425
Db	135250	CAAGCCCGGGAAGACCTCGAGGTGTACGGCGTGTTCAGCACCGTCAGMTGGGACACCC	135309
OY	425	-----	425
Db	135310	AGCACCCCGGCCCTTGCGATGGGCCCTGTCTTTGTCCCTGATTGCGTGACTGTGACTGT	135369
OY	425	-----	425
Db	135370	ACTCTGTACTCTCTCTTTGCGCTTGGCTTGGGGGTCTTTGCTTTTCGAGACTGTGCTTG	135429
OY	425	-----	425
Db	135430	CTGCAGGCACTGTACCCCTGTGCTGCTGCCCATGAGACCAAGSGCTTCCCTTGTCTCT	135489
OY	426	-----SerAlaValPheGlnGlyPhea 	433
Db	135490	GATGGCACTCTGGCGCTGACTGTGCCATGTGGCCCGGAGTGGCGGTGTCCAGGGCTTGG	135549
OY	433	laValCysValTyrrHisMetAlaAspIleTrpGluValPheAsnGlyPProPheAlaHisA	456
Db	135550	CGGTGTGTGTACACATGAGCGCAGACATCTGGGAGGTTCATCAAGCGGGCCCTTGGCCAC	135609
OY	453	rAspGlyProGlnHisGlnTrpGlyProTyrgLysGlyValProPheProAspProG	473
Db	135610	GAGATGGGCTCAGACACAGTGGGGGCCCTTAATGGGGGCAAGGTGCTTCCCTCGCCCTG	135669
OY	473	lyVal-----	474
Db	135670	GGGT--GGTGAATCTGTGGGGCCAGGCCAGGAAGGAGGCTGAGACAGATGTGGGA	135728
OY	474	-----	474
Db	135729	AGGGCTCTTGGAGTGGCTCATGTCTCCCATGCACCCCCACGCGACGCGACTTGGCCAA	135788

[illegible]

Db 136868 GCTTCAGTCAAGCTCTGCTTGAGAGCCAGCATGACAGGGGCTTCTGTAAAC 136927
 QY 582 ----- 582
 Db 136928 CATGCTTGAGAAAGACTGTCTAGCTTACCTGCAGAGAGTCAAGACCGAGCTGGGA 136987
 QY 582 ----- 582
 Db 136988 GCTAGGTGTTCCACCCAGGTTGAGAGAGTGAAGAGGGGTGACGGGAGGTG 137047
 QY 583 -----ValProThrProIleThr 588
 Db 137048 TTGGTCAACCCGACCCACATGAGACCTTCTCTCCAGGTGCCAACAACCTATCAC 137107
 QY 589 GluMetGluIleSerValIysArg ----- 596
 Db 137108 GAATGGAGATCTCTGTCAAAAGGTAAAGACAGCTGCCACCTCTCTTGAGCCCCCT 137167
 QY 596 ----- 596
 Db 137168 TCTGCCCTGTCCGAGGCTCTCTCTGCCCCCTGCTCCGACAGCTCTCGAGCTGCAC 137227
 QY 596 ----- 596
 Db 137228 TGTAGTCTGGGAAAGCTGTGCTGAGGACAGGCTGATTCGCCCTGCAACAAGTGGCCAG 137287
 QY 596 ----- 596
 Db 137288 TGGGCACTGAGGACCCAGTAGTGGGGCCATGTCAGATTAAGGGATAGTCACAAAG 137347
 QY 596 ----- 596
 Db 137348 CCAGTTCCTAAGGAGGCAAAAGTGTGCTGAAAAACAAAAATCCAGATGTTAGTCCAA 137407
 QY 596 ----- 596
 Db 137408 CAAAGCCACTTTTAGAGGCTACTATATTGGCCAGTGGCAAAACCCGTCTCCCATAT 137467
 QY 596 ----- 596
 Db 137468 TTATCTGCTCCAGTGTATCTTTGTATAATCTAACTTTCGTGCAAGGTTGCTTAAGC 137527
 QY 596 ----- 596
 Db 137528 AAGTGCACCTCTAAGTCAAGGACAACAGATAAAGGCCCTGGGATGAGGGGCAAA 137587
 QY 596 ----- 596
 Db 137588 GGTACGACCTGGAGGGCTGAGAGAACCCCTTAGAGAAACATCCAGGCTTGGTTC 137647
 QY 596 ----- 596
 Db 137648 TCCATCCATCCCTTCTGTCTGTCGACAGAAATCTGGTATGACAGATGGGTGATTAACA 137707
 QY 596 ----- 596
 Db 137708 CTTTGTGTTTTTAATGAATCAAGAGGTTCTGGAAGACTTCTGCAAGTGGAGGGGT 137767
 QY 596 ----- 596
 Db 137768 CTCAGGAAGCTGAGGGGCTGAGAGGCTGGCTGGTGAGGGGGAAGCCTGGGCTCACTCA 137827
 QY 597 -----GluMetLeuTyrValIySerArgLeuGlyValAla 608
 Db 137828 GGTCTGCTTCTGCTCCCTCCGCAAAATGCTATACGTGGCTCTCGGTGGGTGGCC 137887
 QY 609 GluLeuArgLeuHisGluCysGluIleThrTyrGlyThrAlaGlyAlaGlyCysGlyLeuAla 628
 Db 137888 CAGCTGCGGCTGCAACAAATGTAAGTACTAGGCACTGCTGTGCAAGTGTGCTGGCC 137947
 QY 629 ArgAspProTyrCysAlaTrpAspGlyAlaSerCysThrHisTyrArgProSerLeuGly 648
 |||||||

Db 137948 CGGGACCCATACTGTGCTGGGATGTGCTCTCTGTACCACCTACGCCACGCTTGGC 138007
 QY 649 LysArgArgPheArgArgGluAspIleArgHisGlyAsnProAlaLeuIleGlyLeuGly 668
 Db 138008 AAGCGCGGTTCCGCGGACGAGACATCCGACAGGCAACCTGCTGTGAGTGGGCG 138067
 QY 669 GlnSerGlnGlu ----- 672
 Db 138068 CAGAGCCAGGAGGTGAGTGTGGCAGGTGGGAGGGCCAGACTGGGCTGGAATCATCC 138127
 QY 672 ----- 672
 Db 138128 CTGTGCCCGGCTCCACCCCTGCGAAGTGCATTTCTCAGATTGTCTTGTATGACGATA 138187
 QY 672 ----- 672
 Db 138188 GTCTCAGTTTCCCTCTCTGTGTAGCTCTCTGTAGTGAATGAGCAGGCTTTGGACA 138247
 QY 672 ----- 672
 Db 138248 CAGTGGCTGGGCTGGGATCCACGCAATGCTTAAGTCTTGTGATACTAGTGGGGCTCTTG 138307
 QY 672 ----- 672
 Db 138308 AGCTGTGTGTCAAGTGTGTGACCTGGGATCTGGTCCCATAGTTCACAGCTGAGCCAA 138367
 QY 673 -----GluGluAlaValGlyLeuValAlaIleThrMetValTyrGlyThr 687
 Db 138368 CCTCTGCTCTCTCCAGAGAGGAGAGTGGGACTTGTGGCAGCCACCATGTCTACGGCAC 138427
 QY 687 rGluHisAsnSerThrPheLeuGlyCysLeuProLysSerProGlnAlaAlaValArgTr 707
 Db 138428 GGAGCAAAATAGCACCTTCCGAGTGTGCTGCCCAAGCTCCCGAGGCTGTGGGCGTG 138487
 QY 707 PheLeuGlnArgProGlyAspGluIleProAsp ----- 718
 Db 138488 GCTCTTGCAGAGGCCAGGAGGTAGGGGCTGACCAAGTGAAGTGAAGAGACCCCTGCTTC 138547
 QY 718 ----- 718
 Db 138548 TAGCCTGTGTGTGTCGAGGTCACTCCATCCGAAACCCCAAGACCTGAGCCCGAGA 138607
 QY 718 ----- 718
 Db 138608 AGGGTTCCTGGGGGAATCTGCCAAGATTAACTAATAAAAGTTGTATTAGGCTCTGG 138667
 QY 718 ----- 718
 Db 138668 GGTGCTAGAAAGAGGTGCCATTATTTTGCCTTTTGGGAATGCCAGGCCAGATGTGCT 138727
 QY 718 ----- 718
 Db 138728 GGAACATCCCCAGTCTGAGGGGTGGGTGATGAGAGAGAGAAAGAGGCTGGGAGGCC 138787
 QY 718 ----- 718
 Db 138788 AGTAAGTCTCCATGTTTAAACACATGGCTCATCTTGAAGGATTCTCTTCAGGGGT 138847
 QY 718 ----- 718
 Db 138848 TCTTGGTTCCTTCTGCCCCCAGGAGGAGGGGAGGGGCTGTGCCCTGGCCAGAGCTAGGC 138907
 QY 718 ----- 718
 Db 138908 TTCAAGAACTCCCTATGCTACACACACACAAGAGAGAACTTCTGGGCTCTTAACCA 138967
 QY 718 ----- 718
 Db 138968 CTGGTCTTGGCTCCACCTCAAAAGCTGGCCCAACACTTCTCTGAGTCAAGCTG 139027
 QY 718 ----- 718
 Db 139028 TCTGCTCAGGGGCTCTCTCCATCCGGAAGGTCTATCAACCCAGAGATGTTTAA 139087

sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

Location/Qualifiers

```

FEATURES
source      1..123943
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3p21.1-9"
            /clone="RPC14-793p23"
            2078..2357
            /rpt_family="AluX"
            /complement(4381..4674)
            /rpt_family="AluSg"
            /complement(6668..6724)
            /rpt_family="MIR"
            /complement(6725..7029)
            /rpt_family="AluSc"
            /complement(7040..7197)
            /rpt_family="MIR"
            /join(7498..7643,7987..8221,8767..8939)
            /gene="OU12F08.x1 Homo sapiens CDNA A1004779"
            8793..8921
            /standard_name="G27739"
            /db_xref="dbSTS:41594"
            9133..9283
            /standard_name="D3S3975"
            /db_xref="dbSTS:10759"
            /complement(join(9446..10863,11044..11116,11221..11313,
            11880..11740,11858..12336,12895..13028,13552..13736,
            14204..14351,14692..14815,15268..15346,15613..15755,
            15838..15899,16397..16516,16913..17045,17993..18047,
            18153..18183,18281..18320))
            /gene="Human mRNA for KIAA0272 gene D87462"
            9497..9735
            /standard_name="D3S3829"
            /db_xref="GDB:626512"
            13712..13962
            /function="Low Coverage"
            /complement(18321..18407)
            /rpt_family="GC-rich"
            18522..18742
            /function="Low Coverage"
            /complement(18554..18661)

```

```

gene
repeat_region 20373..20654
              /rpt_family="AluJb"
repeat_region 20931..21138
              /rpt_family="MIR"
              21525..21781
              /rpt_family="MIR"
              21993..22279
              /rpt_family="AluSx"
repeat_region 23167..23448
              /rpt_family="AluJb"
              /complement(23639..23858)
              /rpt_family="AluSg"
              /complement(23859..24621)
              /rpt_family="L1ME2"
              /complement(24622..24923)
              /rpt_family="AluY"
              24932..25002
              /function="Low coverage"
              24993..25290
              /rpt_family="AluY"
              /complement(25299..25521)
              /rpt_family="L1ME2"
              /complement(25497..26685)
              /rpt_family="L1ME2"
              27118..27246
              /rpt_family="MIR"
              /complement(27249..27544)
              /rpt_family="AluSx"
              27698..27826
              /rpt_family="MIR"
              /complement(27868..28170)
              /rpt_family="AluY"
              31632..32081
              /note="Region: q987a10.x1 Homo sapiens CDNA A1221496"
              31774..31924
              /standard_name="G38576"
              /db_xref="dbSTS:58631"
              31932..32081
              /standard_name="G30450"
              /db_xref="dbSTS:44631"
              32599..32898
              /rpt_family="AluSg"
              33780..33819
              /rpt_family="(CA)n"
              /complement(33968..33996)
              /rpt_family="AT-rich"
              /complement(34011..34082)
              /rpt_family="AluS"
              /complement(34083..34121)
              /rpt_family="(GAAA)n"
              /complement(34123..34410)
              /rpt_family="AluSg"
              /complement(34413..34697)
              /rpt_family="AluJb"
              /complement(34718..35265)
              /rpt_family="L1MC3"
              /complement(35266..35558)
              /rpt_family="AluSg"
              /complement(35561..35738)
              /rpt_family="L1MD3"
              /complement(35739..36036)
              /rpt_family="AluSg"
              /complement(36038..36236)
              /rpt_family="L1MD3"
              /complement(39127..39419)
              /rpt_family="AluSx"
              41664..42236
              /note="Region: Unigene cluster containing W95768 and

```


2y	131	-----	131
2b	51766	CGCCTGTGGGCCACCCCTCCAGCTGAGCGAGCGGCTGCAAAAGACTCCGGCCTTCCTC	51707
2y	131	-----	131
2b	51706	CAGGATCTTCCAGCAGCTTTGCTCTAGTTTCTGCTGTTGGGCGCTTCTCCCTGGGGACT	51647
2y	131	-----	131
2b	51646	GGGTAGGGCACAGATGGGACAGTGGGAAGATTTAGCTGCATGATGAAGGGCTAAGAAAT	51587
2y	131	-----	131
2b	51586	GGAGGGGTGGAGGTAGGAGGCCCGAGGTAGGCTGGGCCAGTGGAGAGCTTGAGCCTCCC	51527
2y	131	-----	131
2b	51526	GTCCCACTGTCCCGCAGGACCGCAGGCTGCCGAAAGCCTTCCAGCACCATGTGATGGAAC	51467
2y	131	-----	131
2b	51466	ATTTCAGATGTAGTGTGTCAAGGGGTGATGCTCAGCCGGGGGCTCCACCTCCACTCC	51407
2y	132	-----	132
2b	51406	TTGGCTGGGAACTAGGCTGAGCTGTGGCCCTGATCTGCCTCCATCATCTCTCTAGA	51347
2y	132	-----	132
2b	51346	CTCTCTGTCTGCCAACCCCTGTGCTCCATCTTCTGGGGCCCCAGGGGCTCCCTGAACTTCA	51287
2y	152	naIaMeTylrLeuAspGluTyrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTy	172
2b	51286	GGCCATGTACTAGTAGTAGTACCGAGACCGGCTCTTCTGGGTGGCTGGAGCGCCTCTG	51227
2y	172	rSetLeuArgLeuAspGlnAlaTTPProAspProArgLeuAla	186
2b	51226	CTCTGTGGGCTGGACCGAGGATGGCCAGATCCCGGGAGGT-GAGCTGAGCTGGCAGGG	51186
2y	186	-----	186
2b	51167	AAGAGGCTCTTTTGAGAACTGGGCACCTCCAGATGGAGGCCAGATGGGAAACTGAGC	51108
2y	187	-----	190
2b	51107	CTGTGAGATCAGGGAGGAGGACTGAGGTGAGCGGGGCTGTGCCAGTCTGTGGGCACC	51048
2y	190	ogInProGlyAlaInArgGluGluGlyValArgLysGlyArgAspProLeu	206
2b	51047	GCAGCCAGGACAGAGGAGAGTGTTCGAAGGGAAGAGATCCTTT-GGTGAGTGTGTG	50989
2y	206	-----	206
2b	50988	TCGGGAGGGATGTCCCCCGAGCTGACACCTGCCCTTACCCCCAGGACAAATCGGTGACC	50929
2y	206	-----	206
2b	50928	ACGCCCCAGTACTCTTACACAGGGGCTTTGGAGCAGATTTGAGGGCTTTGAAGGTCT	50869
2y	206	-----	206
2b	50868	GCTGGGATGTCAACACCACCCCTCCCAAGAGCTGACGCGCAGACAGGGCTGAGCCTT	50809
2y	206	-----	206
2b	50808	GGGGCCCAAGAGATGTGGGGCACCCTCTTCACCCCTCAACCAACCCCTCTTCTCCC	50749
2y	207	----ThrGluCysAlaAsnPheValArgValLeuGlnProHisAsnArgThrHisLeuLe	225
2b	50748	GCAGACAGAGTGGCGCCAACTCTGTGGGGGTGTACAGGCTCAACAGCGAGACCACCTGCT	50689
2y	225	uAlaCysGlyThrGluAlaPheGlnProThrCysAlaLeuLeuIleThrValGlnHisArgG	245

Db	50688	AGCGTGTGGCACTGGGGGCGTTCCAGACCCACACCTGTGCTCCCATCAGAGTTGGCGACCGTGG	50629
QY	245	YGLU-----	246
Db	50628	GGA-GGTAGCCCTGGGGCCAGGCGCCACAAAGGAGCTGGGAGGCGCATGACATTTCTCCCTCG	50570
QY	246	-----	246
Db	50569	CTGCCTGAACATTTCCCATTTGAACCTGGGTTAGGAGACAGAGGGGCGAGGGGCGTGAAGGGCG	50510
QY	246	-----	246
Db	50509	CTTTGTGTTGGGGAGCTGACCCCATTTAAGAGGGGCGCTGGATGAGAGAGCGCTTCCGCCAC	50450
QY	246	-----	246
Db	50449	CCCCGCTCTGCGGCCGACGACGAGGGGCTCTGTTGGGTGGGCTGGTGCTGCTTCT	50390
QY	247	-----HisValLeuHisLeuGluDr	253
Db	50389	CCCAAGTGAGCGCCCTCCCTCCCTCTCTCCCTCCCACTAGACATTTGCTCCACCTGGAGCG	50330
QY	253	OGLSerValGluSerGlyArgGlyArgCysProHisGluProSerArgProPhalaSe	273
Db	50329	TGGCAGTGTGGAAGATGGCCGGGGGGGTGGCTCCACAGAGCCAGCGCTCCCTTGGCAG	50270
QY	273	TTThPheIle-----	276
Db	50269	CACCTTCAT-AGTGGGTGATGCCACAGGCGAGGTGAGAACAGGAGGGGTGCACGTGGC	50211
QY	276	-----	276
Db	50210	CTATGTCACTGGGAGGAGGAGGAGGCCGGGCTGTGGGCTGGGCGCTCTCCCTGACTGTG	50151
QY	277	-----AspGlyGluLeuYrThrGlyLeuThrAlaaspPheLeuCl	290
Db	50150	TTCCCAACCCCGACCCCGAGAGGGGAGCTGTACACGGGTCTCATCTGCGACTTCTCTGG	50091
QY	290	YArgGluAlaMetIlePheArgSerGlyGlyProArgProAlaLeuArgSerAspSerAs	310
Db	50090	GCGAGAGGCCATGATCTTCCGAAAGTGAGAGTCTCGCCGACGCTCGCGTTCGACACTGA	50031
QY	310	PGLIserLeuLeuHisasp-----	316
Db	50030	CCAGAGTCTCTTGCACGG-TGAGGCTCTGCCCTGGAACTGGCAGGGGTGAAGAGGCTACC	49972
QY	316	-----	316
Db	49971	TTCCAGCTAGGACACACAGCCTCAGTGGAGTCCCTGCCCCATTCTCTCCCACTGG	49912
QY	316	-----	316
Db	49911	GTTCTACTCCATCTCTCCCTGTGGGGCCCTCCACCCAGCATGACCTGTACCCCTTCC	49852
QY	317	-----ProArgPheValMetalAlaArgIleProGluAsnSerAspGlnAspAsnAspIy	335
Db	49851	AGACCCCGCGTTTGTGATGGCGCGCCCGGATCCCTAGAACTGTGACAGAGCAAGAGACAA	49792
QY	335	SValIyrPhePhePheSerGluThrValProSerProAspGlyGlySerAsnHisValIth	355
Db	49791	GGGTACTTCTTCTCTCTGGAAGACGTCCTCCCGCCGATGGTGGCTCGAACCATGTCTAC	49732
QY	355	rValSerArgValGlyArgVal-----	362
Db	49731	TGTCAGCGCGGTGGCGCGCTCTGCGTGTAAAGACTGGAAAGGGGTGTGAGGTTCAAG	49672
QY	362	-----	362
Db	49671	TCCGCTACACACCTCCCATTCAGACCTTCTCTCCCGACACCCCTGTGGCGACAGTAAAGG	49612
QY	362	-----	362

QY 585 ThrProIleThrgIuMeGluTleSerValIysArg
 Db 47395 AACCTATACCGAATAAGATGATCTGTCTCAAAAGGTAAGAACGTCGCCACCTCTCTC 596
 QY 596
 Db 47335 TTGACCCCCCTTTGCCCCCTGCTCCGACGCTCTCTCTCTGCCCCGCTCCGAGCCTC 596,
 QY 596
 Db 47275 CTGAGCTCCACTGTAGTGTGGGAAAGCTGTCTAGGCGAGCTGATTCGCCCCGTGA 596
 QY 596
 Db 47215 ACAAGTGGCCAGTGGGCTGAGGCGACCGAGTAAGTGGGCGCATGTCCAGTAAGGGA 47156
 QY 596
 Db 47155 TAGTCAAAAGGCCAGTTCTTAAGGAGGCAAAAGTGTGCTGAAAACACAAAATCCAGA 596
 QY 596
 Db 47095 TGTTAAGTCCAAAGGCCACTTTTAAAGGCTCTACTATATTTCCCAAGTGGCAAAACC 596
 QY 596
 Db 47035 GTCTCCCATATTTATCTGCTCCAGTGTATCTTGTGTAATCTAATCTTGTGTCAAG 46976
 QY 596
 Db 46975 GTTCTTAAGCAAGGTGCACCTTAAGTCAAGGCACTAGAGATAAAGCCCTGG 596
 QY 596
 Db 46915 ATAGAGGCAAAAGTCAACCTGGAGGCGCTGAGAGGCCCTTAGAGAAAACATCCC 596
 QY 596
 Db 46855 AGGTCTTGTTCTCTCTCATCTCTCTGTCTGTCTGTGCAAGAAATCTGTATGACAGAT 46796
 QY 596
 Db 46795 GGGGTATAAACACTTTTGTGTTTGAATGAATCAAGAGTTCTGAAAGACTTCTGCA 596
 QY 596
 Db 46735 AGTGGCAGGGGTCTCAGAAAGCTGAAGGCTGAGAGCTGCTGCTGGGCGAAACCT 46676
 QY 597
 Db 46675 GGGCTACCTCAGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 604
 QY 605 LeuGlyValAlaGlnLeuArgLeuHisGlnCysGluThrArgGlyThrAlaCysAlaGlu 46616
 Db 46615 CTGGGTGTGGGCGAGCTGCGCTGCAACCAATGTAGACTTACGGCACTGCTGTGCGAG 46556
 QY 625 CysGlyLeuAlaArgAspProTyrCysAlaIlePhePheValAspSerCysThrHisTyrArg 644
 Db 46555 TGTCTGCTTGGCGGAGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
 QY 645 ProSerLeuGlyLysArgArgPheArgGlnAspIleArgHisGlyAsnProAlaLeu 46496
 Db 46495 CCCAGCCTTGGCAAGCGCGGTTCGCGCGGAGACATCCGGCACGGCAACCCCTCCCTG 46436
 QY 665 GlnCysLeuGlyGlnSerGlnGlu
 Db 46435 CAGTCCCTGGGCGCAGAGCCAGAAAGTGTGTGCGCAGTGGGAGGCGCCAGACTGGCC 46376
 QY 672
 Db 46375 TGGAACTCACCCCTGTGTGGCGGCTCCACCCCTGCGAAGTGAATTTCTCAATTTGCTTGG 46316
 QY 672
 Db 46315 TCATGACGATAGTCTCACTGTTCCCTCTCTGTGTAGTCTGTCTGTATGATGAAAGCG 46256
 QY 672
 Db 46255 GGTCTTTGGACACAGTGTGCTGGCTTGGGATCCAGCATGCTAGTTCTTGTCACTA 672
 QY 672
 Db 46195 GTGGGCTCTAGCTGT 672
 QY 673
 Db 46135 AAGTGAAGCCACTCTGCT 46136
 QY 683 lValTyrGlyThrGluHisAsnSerPheLeuGluCysLeuProLysSerProGlnAl 683
 Db 46075 GGTCTACGGCGACGAGACATAGCACCTTCTGAGAGTCCCTGCGCAACTCTCCCAAGC 46076
 QY 703 AlaValArgTyrPheLeuGlnArgProGlyAspGluGlyProAsp
 Db 46015 TGTCTGTGCTGCTGCT 46016
 QY 718
 Db 45955 GACCCCTGCTCTAGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 718
 QY 718
 Db 45895 TGAAGCCCGAAGAAAGGGTCTGGGGGAATGTCCAGATTAACTAAAGATTGTA 718
 QY 718
 Db 45835 TTAAGCTCTGGGTGCTTAAGAAGTGTGCTTATTTTGTGCTTTTGGAAATGCCAGG 45776
 QY 718
 Db 45775 CCAGATGTGCTGAATCTCCCAAGTGTAGGGGTGGTATGAAGAAGAAAGAGG 45716
 QY 718
 Db 45715 TCTGGCAGGCCAGTAAGTCTCCATGTTTACACATGCTACTTTTGAAGATTTCTC 718
 QY 718
 Db 45655 TTTCTCAAGGGTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
 QY 718
 Db 45595 CCAAGCTAGGCTTCAGAACTCTCTATGTACACACACAAAGAAAGAACTTCTG 718
 QY 718
 Db 45535 GGTCCCTAACCACTGTCT 718
 QY 718
 Db 45475 GGAGTCAAGTGTCTGCTCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
 QY 718
 Db 45415 GGGATGTTTAAGTCTGGGCTGAGATGGGATTTGGCGGGGAGGAAGGAGGGG 45356
 QY 718
 Db 45355 CAGAGTGTATGATTACAGGAGGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718
 QY 718
 Db 45295 CCTTGGCCAGCTGTGGCTGTGGCTGAGAGATTTCCCAAGGGAGAGGAGGAGAAC 45236
 QY 718
 Db 45236

45235 AGATGGTGTTCACACACTAGCTAATGGTCAAGCAATCCACCCACACACAGATG 45176
OY 718 ----- 718
Db 45175 TGCAGACTAAACCCACACAGGGTCCAGCCAGGCGGCTGGATCAGGGCGGCACTC 45116
OY 718 ----- 718
Db 45115 CCACCTCCACGCTGTATGAGGCGATCAGCCACCCCTGGGCTGTACAGCTGCAC 45056
OY 718 ----- 718
Db 45055 CGTTGGGGGATGAAGGCTTCAGACCCAGGATGGAAAGCCCTGTGTGTCTGTGTGT 44996
OY 718 ----- 718
Db 44995 GTGTGCTGCACGTGATGTGTGCTGGGACACGACATTTGTATGTGTATAGATAGGG 44936
OY 718 ----- 718
Db 44935 AGTTTCTGAAGCCAGTTCGCGGACAGGGGACAGCCCTCCCTTGAGAAAGCCTGTA 44876
OY 718 ----- 718
Db 44875 GATGGATGATCTCTCAGAGGCTTAAGGGATATACATCTTACCCCTATCTCTTGG 44816
OY 718 ----- 718
Db 44815 GAATGTGTCCAGGCGCTCTGAGGCCAAGATTGAGTCTCTGACTAAGGAGAAACCA 44756
OY 718 ----- 718
Db 44755 GCCAGTCAATCTTACAGAGGCGCTCTCTGCTATGGAGGTGAGGCCACCTGA 44696
OY 718 ----- 718
Db 44695 ATTAGTGTGTCTTAGAGAGGCCAGACTGCTGCTGCTCTCTCTCTCTCTCTCTCT 44636
OY 718 ----- 718
Db 44635 GGAGAAATGGGCAAGAGTGGCCGACGACTGGCCCAAGATGCGAGAAAGCTGGGGC 44576
OY 718 ----- 718
Db 44575 AGAGAGCCGATGTAACTCCACCTCTGACCTCTCTCTCTCTCTCTCTCTCTCTCT 44516
OY 718 ----- 718
Db 44515 AGCTGAAGAGCGAGCGAGCTCTGACACGAGAGGGGCTGCTCTCTCTCTCTCTCTCT 44456
OY 718 ----- 718
Db 44455 GCGCTTCGATGGGCGACCTACCTGACCTGACCTGACCTGACCTGACCTGACCTG 44396
OY 718 ----- 718
Db 44395 TGTCTCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44336
OY 718 ----- 718
Db 44335 AGCCAAAGCCAGAGAGGCGCCGAGGCGGAGGCTGCTCTCTCTCTCTCTCTCTCTCT 44276
OY 718 ----- 718
Db 44275 GGTACAGAGGATCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44216
OY 718 ----- 718
Db 44215 GTAGCGGCTGT 44156
OY 718 ----- 718
Db 44155 GGGGCAAGAGCGCCAGGCGGCAAGAGCTGGGCTGAGAGCTGCAAGAGATGAAGA 44096

OY 859 erArgValHisaleGluHisAsnArgThrProArgGluValGluAlaThr 875
Db 44095 GCCGGGTGATGCCAGACATCGAGCGCCCGGAGGTGGAGCCACG 44046
RESULT 9
AX155172 2337 bp DNA linear PAT 22-JUN-2001
LOCUS sequence 3 from patent WO0140278.
DEFINITION
ACCESSION AX155172
VERSION AX155172.1 GI:14536657
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 2337)
Holloway, J.L. and Foley, K.P.
Human semaphorin zsmf-16
Patent: WO 0140278-A 3 07-JUN-2001;
TITLE
JOURNAL
ZymoGenetics, Inc. (US)
location/Qualifiers
FEATURES
source
1. 2337
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="degenerate polynucleotide sequence of zsmf-16"
BASE COUNT 308 a 273 c 468 g 293 t 995 others
ORIGIN
Alignment Scores: 2337
pred. No.: 573
Score: 3102.50
Percent Similarity: 71.75%
Best Local Similarity: 71.62%
Query Match: 65.37%
Gaps: 2
US-09-813-290-2 (1-875) x AX155172 (1-2337)
OY 94 MetaLalProserAlaThrPalalIleCysTrpLeuLeuGlyLeuLeuLeuHisGly 113
Db 1 ATGCGCNCNMSGNCNTGGCGNAHTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 60
OY 114 SerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTrpArgAspLeu 133
Db 61 WSNWSNGNCNMSGNCNCGNCCNMSGNCNMSGNCNMSGNCNMSGNCNMSGNCNMSGNCN 120
OY 134 Leu-----Ser 135
Db 121 ATGCTNMGNAARCCNMSNNACTGTGATGATGATGATGATGATGATGATGATGATGATG 180
OY 136 AlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTrp 155
Db 181 GCNAAAYGMSNCGNATHTTGT 240
OY 156 LeuAspGluTrpArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTrpSerLeuArg 175
Db 241 YTNAGYGARIVMNGAVMNGVNTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 195
OY 176 LeuAspGlnAlaTrpProAspProArgGluValLeuTrpProGlnProGlyGlnArg 330
Db 301 YTNAGYCAARGCTGGCCGCAATCCNMGNGAR----- 215
OY 196 GluGluCysValArgGlySerProLeuThrGluCysAlaAsnPheValArgVal 357
Db 331 -----ACNGARTGCGMAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 235
OY 216 LeuGlnProHisAsnArgThrHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThr 417
Db 358 YTNCAARCCNCAIVAMNACNCAAYTNTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 255
OY 236 CysAlaLeuLeuIleThrValGlyHisArgGlyGlnHisValLeuHisLeuGlnProGlySer 477
Db 418 TGYCCTNTNATHACGNTGNGCAVMGNGARGARCAVATNTNCAAYTNTGARGCCNGMSN

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frazz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollaway, J.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,
Kovar, C., Kratoch, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, R.,
Louliseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Nickerson, E., Nwokenkwo, S.,
Nguyen, A., Nguyen, N., Nguyen, N., Oviedo, R., Pace, A., Payton, B.,
Ogih, M., Okunou, G., Oragunye, N., Pickens, R., Prims, E., Pu, L.L.,
Peery, J., Perez, L., Peters, L., Pickens, R., Rojudo, K., Rojudo, K.,
Quiles, M., Reay, Y., Rives, M., Rojas, A., Rojudo, K., Rojudo, K.,
Ruiz, S., Seavey, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,
Sisson, I., Sodergren, E., Sotak, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Stang, H., Sutton, A., Swatek, A., Taylor, C., Taylor, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, G., Wleczek, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 229706)

Worley, K.C.

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:16901619.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GC21

Center clone name: CH230-8P2

----- Summary Statistics

Assembly program: Phrap: version 0.990329first call to

findhaplolist

Consensus quality: 21655 bases at least Q40

Consensus quality: 21654 bases at least Q30

Consensus quality: 22067 bases at least Q20

Estimated insert size: 219330; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 3.4x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length

NOTE: (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

NOTE: This is a 'working draft' sequence. It currently

45529 59289: contig of 13761 bp in length

59290 59389: gap of unknown length

59390 73166: contig of 13777 bp in length

73167 73266: gap of unknown length

73267 86647: contig of 13381 bp in length

86648 86747: gap of unknown length

86749 100569: contig of 13822 bp in length

100570 100669: gap of unknown length

100670 115083: contig of 14414 bp in length

115084 115183: gap of unknown length

115184 127043: contig of 11859 bp in length

127043 127143: gap of unknown length

127143 134393: contig of 7251 bp in length

134394 134493: gap of unknown length

134494 142906: contig of 8413 bp in length

142907 143006: gap of unknown length

143007 148251: contig of 5245 bp in length

148252 148351: gap of unknown length

148352 154123: contig of 5774 bp in length

154126 154225: gap of unknown length

154226 162069: contig of 7844 bp in length

162070 162169: gap of unknown length

162170 167280: contig of 5111 bp in length

167281 167380: gap of unknown length

167381 174948: contig of 7568 bp in length

174949 175048: gap of unknown length

175049 179189: contig of 4041 bp in length

179190 183684: gap of unknown length

183685 183784: contig of 4495 bp in length

183785 189316: gap of unknown length

189317 189416: contig of 5532 bp in length

189417 195866: gap of unknown length

195867 201122: contig of 6450 bp in length

201123 201222: gap of unknown length

201223 205343: contig of 5156 bp in length

205344 205443: gap of unknown length

205444 208257: contig of 4121 bp in length

208258 208357: gap of unknown length

208358 211417: contig of 2814 bp in length

211418 211517: gap of unknown length

211519 214163: gap of unknown length

214164 214263: contig of 2646 bp in length

214264 217312: gap of unknown length

217313 217412: gap of 3049 bp in length

217413 219666: contig of 2534 bp in length

219667 220066: gap of unknown length

220067 221858: contig of 1792 bp in length

221859 221958: gap of unknown length

221959 224415: gap of unknown length

224416 224515: contig of 2457 bp in length

224516 225533: gap of unknown length

225534 225633: contig of 1018 bp in length

225634 226997: gap of unknown length

226998 227097: contig of 1364 bp in length

227098 228527: gap of unknown length

228528 228627: contig of 1430 bp in length

228628 229706: gap of unknown length

229707 229706: contig of 1079 bp in length.

FEATURES

Source

1. 229706

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="CH230-8P2"

/location="54720 c 55600 g 58729 t 3213 others

BASE COUNT

57444 a 54720 c 55600 g 58729 t 3213 others

ORIGIN

Alignment Scores:

1.32e-81

Score: 2282.50

Percent Similarity: 26.19%

Best Local Similarity: 24.99%

Query Match: 48.09%

Length: 229706

Matches: 648

Conservative: 31

Mismatches: 64

Indels: 1854

US-09-813-290-2 (1-875) x AF034744 (1-2898)

QY 36 ArgGlnGlyPro-----SerLeuLeuSerSerAlaProLeuProAla 50
 Db 126 AAGCGGGCTCCTTGAGACGCGCTTCTCGCCCTTTTACACACCGGGCCCAAGACAGA 185
 QY 51 GlnAsp---TrpValGluProLeuProTyrLysTrpLysProGlyLysSerAlaAsn 69
 Db 186 AAGGCTTAGCGGATCCAAATATGTGCCGCAATGACACTGGGAATGGATTTTCTGAT 245
 QY 70 TyrAsn-----ArgArgProAlaGlyProGluGly----- 80
 Db 246 GACAAACCTCTTCTTTGGACAAACCTGTGCGCCGAGTTGCCCTTGAGGGAAGTAC 305
 QY 81 -----SerAlaGlyArgArg-GlnArgCysProGlnPhePr 92
 Db 306 TAAGTAAACTCAATCTCTTAAAGTGTGGCTGACGGGGCCAGAGAGAGCCACAGC 365
 QY 92 oSerMetAlaProSerAlaTrpAlaIleCysTrpLeuGlyGlyLeuLeuLeuIleG 112
 Db 366 CACCATGGCAGCGCGGACACATCCTCCTGCTGCTGGGTGCTACCTGCTGGAACCT 425
 QY 112 yGlySerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTrpArg 132
 Db 426 CTGACACCCAGGTCACTCGCGGACACCCCTCTTACCCAGGCTACGCTCTACATTAAGA 485
 QY 132 PLeuLeuSerAlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuG 152
 Db 486 ACTTTGGAACTGAATAGGACTTCATATTTCAAAAGCCCCCTTGGATTTCTTGATGTCCA 545
 QY 152 nAlaMetTrpLeuAspGluTrpArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTy 172
 Db 546 TACATGTGCTGCTGGATGAGTATCAAGAAAGGCTTTGTGGGAGGACAGACCTTGCTCA 605
 QY 172 rSerLeuArgLeuAspGlnAlaTrpProAspProArgGluValLeuTrpProGlnPr 192
 Db 606 TTCCCTGAACCTTGAAGACAGTACGACGCGCTACAGAGATATCTGGCCGACAGCAGC 665
 QY 192 oGlyValArgGluGlyCysValArgGlyArgArgAspProLeuThrGlyCysAlaAsnPh 212
 Db 666 AGTAAAGGTAGAAAGATCATATATGAAGAAAGAAC--GCMAATGGAGTGGCCAAATTA 722
 QY 212 eValArgValLeuGlnProHisAsnArgThrHisLeuLeuAlaCysGlyThyGlyAlaPh 232
 Db 723 TATCCGGGTTTGGATCCTCAACAGAGACACACTTCTGACCTGTGCTACTGGAGCTTT 782
 QY 232 eGlnProThrCysAlaLeuIleThrValGlyHisArgGlyGluHisValLeu---HisLe 251
 Db 783 TGATCCACACTGTGCTTCATCAGAGTGGGACCATTCAGAGGAACCCCTGTTTCACCT 842
 QY 251 uGluProGlySerValGluSerGlyArgGlyArgCysProHisGluProSerArgProPh 271
 Db 843 GGAGTCACACAGATCTGAGAGAGAGAGGAGGAGATGCTCTTTGACCCCAACTCTCCTT 902
 QY 271 eAlaSerThrPheIleAspGlyGluLeuTyrThrGlyLeuThrAlaAspPheLeuGlyAr 291
 Db 903 TGTGTCCACGCTAGTGGGAATGAGCTGTTGTGCTGAGACTTACAGTCACTATTTGGGGCAG 962
 QY 291 gGluAlaMetIlePheArgSerGlyGlyProArgProAlaLeuArgSerAsp---SerAs 310
 Db 963 AGATCGGCGATCTTCCGACAGATGGGAACTTACGCCATTTGCGACTGTAGCATGACGA 1022
 QY 310 pGlnSerLeuLeuHisAspProArgPheValMetAlaAlaArgIleProGluAsnSerAs 330
 Db 1023 TGAAGCGGCTCTGTGAAGAACCAAAATTTGTAGTTCAATATGATTTCTGTATACGAAGA 1082
 QY 330 pGlnAspAsnAspArgValTyrTrpPhePhePheSerGluThrValProSerProAspGlyG 350
 Db 1083 CCGGATGACAAACAAATGTAATTTCTTTACTAGAGAAGCGCTGTGAGCGGGAACAA 1142
 QY 350 ySerAsnHisValThrValSerArgValGlyArgValCysValAsnAspAlaGlyGlyG 370
 Db 1143 CGGC---CACACAGATCTACACCCGAGTGGGGGCGGTGCTGCTGATGATGAGAGGACA 1199

QY 370 nArgValLeuValAsnLysTrpSerThrPheLeuLysAlaArgLeuValCysSerValPr 390
 Db 1200 GAGAAATCCCTGGTGAACAGAGGAGCACTTCTTAAAGCCGCGCTGGTTGCTCAGTGGC 1259
 QY 390 oGlyProGlyValAlaGluThrHisPheAspGlnLeuGluAspValPheLeuLeuTrpPr 410
 Db 1260 GGGAAATGATGGAATGACACATCTTGTGACGAACTAGAGATGTGTCTTACTGGCCGAC 1319
 QY 410 oLysAlaGlyLysSerLeuGluValTyrAlaLeuPheSerThrValSerAlaValPheG 430
 Db 1320 CAGAGATCTCTAAGAAATCCAGATGATATTTGGACGTGTTAATACACAGCATATATTAG 1379
 QY 430 nGlyPheAlaValCysValTyrHisMetAlaAspIleTrpGluValPheAsnGlyProPh 450
 Db 1380 AGGCACTGCTGATGTGTGTATCACAATGTCACGGAATCCGGGAGCCCTTAAATGGCCATA 1439
 QY 450 eAlaHisArgAspGlyProGlnHisGlnTrpGlyProTyrGlyGlyLysValProPhePr 470
 Db 1440 TGCTCATTAAGAAAGCGCCGAAATACACACTGCTCATATATGAAGGAAAGTCCCTTACC 1499
 QY 470 oArgProGlyValCysProSerLysMetThrAlaGlnProGlyArgProPheGlySerTh 490
 Db 1500 AAGCCTGCTGCTGCTGCGCAGCAAAAGTAAAC-----GGAGCCAGATGAGAACAC 1550
 QY 490 rLysAspTrpProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpPr 510
 Db 1551 CAAGATTTCCCGGATGACGCCATCCGTTCCGAGATGATCCCTTAATGATATCAGCC 1610
 QY 510 oValArgProArgHisGlyArgProValLeuValLysThrHisLeuAlaGlnGluLeuH 530
 Db 1611 CAAAAAACCTGTTCAATAAAAAACCATACTGTAATAAACAGATGGAATATACACTGAG 1670
 QY 530 sGlnIleValValAspArgValGluAlaGluAspGlyThrTrpAspValIlePheLeuG 550
 Db 1671 GCAACTTCCCGGTGATCGGCTGGAAGCGGAGATGCGCAGTGTGACGTCCTTATTTATTCG 1730
 QY 550 yThrAspSerGlySerValLeuLysValIleAlaLeuGlnAlaGlyGlySerAlaGluPr 570
 Db 1731 GACAGACACAGAAATGTGCTGCAAAAGTAAATCATCTTAAACCAAGAAACAGATGGAT 1790
 QY 570 oGluGluValValLeuGluGluLeuGlnValPheLysValProThrProIleThrGluMe 590
 Db 1791 GGAGGAAGTCATTTAGAGAACTTCAAAATATVCAAGATCCAGGCCCTATATTTCTAT 1850
 QY 590 tGluIleSerValLysArgGlnMetLeuTyrValGlySerAlaGluGlyValAlaGlnLe 610
 Db 1851 GGAATTTCTTCAAAAGACAGAACACACTTTACATTGGATCAGCCTGCTGTGGCAGAGT 1910
 QY 610 uArgLeuHisGlnCysGluThrTyrGlyThrAlaCysAlaGluCysCysLeuAlaArgAs 630
 Db 1911 CAGATTCATCACTGCGACATGTATGGCAGTGCTGTGCTGACTGCTGCGGTGCGAGA 1970
 QY 630 pProTyrCysAlaTrpAspGlyAlaSerCysThrHisLysTrpArgPro-----SerLeuG 648
 Db 1971 CCCGTAAGTCCCTGGATGGCATTTCTGCTCCAGGTACTTACCAACAGAGTGCACAGC 2030
 QY 648 yLysArgArgPheArgArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCysLeuG 668
 Db 2031 AAAGAGGAGGTTCGCAAGGACGAGCTTCCGCAATGCCAAGCGCCCAAGTGGTTTGG 2090
 QY 668 yGlnSerGlnGluGluGluValValGlyLeuValAlaAlaThrMetValTyrGlyThyG 688
 Db 2091 ACAGCAATTTTGGAGCGGCTGGACAGAGACTGMAAGAGGCTGCTTATGGCATTAAGA 2150
 QY 688 uHisAsnSerThrPheLeuGlnCysLeuProLysSerProGlnAlaAlaValArgTrpLe 708
 Db 2151 GAGCAACAGTACTCTGTGGAATGACACCCCGCATCATACAAACAAAGTCACTGTGTT 2210
 QY 708 uLeuGlnArgProGlyAspGluGlyProAspGlnValLysThrAspGluArgValLeuH 728
 Db 2211 TGTACAGAAAGGAGCGACGCTTAAGAAAGAGAGGTGAAGAGGATGACAGATTGTCA 2270

DB	BASE COUNT	ORIGIN	665 a	542 c	595 g	526 t
			<pre> /protein_id="CA802590.1" /db_xref="GI:3150196" /db_xref="SPTREMBL:P70275" /translation="MAPACHIT1L1MGLHLEMTPEGHANSPYRLRLSHKELLEIN RISIFQSPGLEFLDHTMLIDELDYORLEVGGRDLVYSLNRYSGVSEIETWSPYAVKV EEDIMKQKANEACNATIRVLDAHNYRNLHLCATGADPRNCAITGVGHNSPEPLPHLE HREERGRGCRPRPNSSVSTLVQNEFLAGLSUYWRDASLTFSGMGLGHTPREND ERLKEPKFVGVSNIPNEDNRDNKMTFFLEKLELENNATITVTRKGLCVNDMG ORTLVNKSITFLKARLVGSVGMNGIDITFEDLEADVLLPRDKRNPVTFGLPNTS TFGHANCSYVHMSIRFANPGRYAKRCEPYNWLSYKAYPRPGSCASVKNQKGVG TTDYRDPDAIRPAIDRPIORIKRVAKRPLVMDGKYNLRQALDRAVEDSDQYGS LFETDGLVLTAVLTITTYNOETEMWNEVLELETFKQPARITISMSISKROOYRHS SAAQVFEHNHMDTGSACDCLANDRYCANDGISGYSYRGTAKHAPRPROVYRHS NAAQCGGQOFGVGBADLRTERLRYVGLSNTSLTCTPRSQAYITWVQKGRVRRK EUKIDRNVKMDLGLFLTVRKSDAGVFCQTVENHFNHYKLTILEVEYKVGSM HKDEERHNNRMPRCPLSGMSQGTQRPWKERLEOLIGNSFQVEEYCEKWCCTKRKR KIKMSPKMYKANFOEKRLRSKAEHFRHTLLS" </pre>			
			<pre> BASE COUNT 665 a 542 c 595 g 526 t ORIGIN </pre>			
			<pre> Alignment Scores: Score: 6.9e-74 Proed. NO.: 2043.50 Percent Similarity: 66.71% Best Local Similarity: 50.57% Query Match: 43.06% DB: 10 </pre>			
			<pre> US-09-813-290-2 (1-875) x MMSEMH (1-2328) </pre>			
Qy	94	MetaLProSeFaLaTrpAlaLeuGlySTrPLeuLeuGlyLeuLeuHisGlyGly	113			
Db	1	AMGGCACCGCGCGGACACATCTCCACCTTCCTGCTGGSGGTACACGTCGGAACCTGCG	60			
Qy	114	SeSerGlyProSerProGlyProSeVaLProArgLeuArgLeuSerGlyArgAspLeu	133			
Db	61	ACCCGCGTGCACCTCCCGGAACCTCTCTACCCGAGGTACAGCTGTGCATTAAGAACTT	120			
Qy	134	LeuSerAlaAsnArgSerAlaLeuPheLeuGlyProGlnGlySerLeuAsnLeuGlnAla	153			
Db	121	TTGGAACTGAATAGAGACTTCATATTTCAAGACGCCCTTGATTCCTGATCTCCATACA	180			
Qy	154	MeLtyrLeuAspLeuTrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSer	173			
Db	181	ATGCTGCTGATAGATGATCAAGAACGGCTCTTGTGGGAGGACAGACACTTGTCTATTCC	240			
Qy	174	LeuArgLeuAspGlnAlaTrpProAspProArgGluValLeuTyrProGlnProGly	193			
Db	241	CTGACTTGGACAGCAGCTCAGGACGGCTACAGAGATATGACTGGCCGACAGCAGACTA	300			
Qy	194	GlnArgGlnLeuGlyValArgGlyGlyArgAspProLeuThrGlnGlyAlaAsnPheVal	213			
Db	301	AAAGTAAAGATGTCATATGAAAGAAAGAAAGAAC--GCAAAATGATGTGGCCATATATATC	357			
Qy	214	ArgValLeuGlnProHisAsnArgThrHisLeuLeuAlaGlyGlyThrGlyAlaPheGln	233			
Db	358	CGGGTTTTGCACTACACAGAGACACACTTCTTGACCTGGTGTACTGGAGCTTTTGAT	417			
Qy	234	ProThrGlyAlaLeuLeuThrValGlyHisArgGlyGlnHisValLeu--HisLeuGln	252			
Db	418	CCACACTGTGCTTCATCAGAGATCGGGGACCAATTCAAGAGAAACCCCTGTTCACCTGGAG	477			
Qy	253	ProGlySerValGlyArgGlyArgGlyGlyProHisGlnGlnProSerAlaTrpProPheAla	272			
Db	478	TCACACAGATCTGACAGAGAGAAAGGGGCGAGATGCTCTTTTGACCCCAACTCTCTTTGTG	537			
Qy	273	SeThrPheLeuAspGlyGlyLeuTyrThrGlyLeuThrAlaAspPheLeuGlyArgGln	292			
Db	538	TCCACGCTAGTTGGCAAGATGAGCTGTTGCTGGAGCTACAGATGACATTGGGCGACAGAC	597			
Qy	293	AlaMetThrPheArgSerGlyGlyProArgProAlaLeuArgSerAsp--SerAspGln	311			
Db	598	TGGCGGATCTTCGACAGCATGGGGGAAGTTTAGGCGCAATATTCGACATGACATGAG	657			


```

/protein_id="AB80952.1"
/db_xref="GI:2522206"
/translation="MLGRMAADLLILALGELLELPAGYHATDTPORPLRLSHKL
MDNRTSVRSPRGVYLHMLDEYOERFLVSGRDLXLSIDRISNNRHEIMWST
PLDAECITIKGRADKCANVRLHRYNRHLLACGMDAPVCTEFLRVNRNAAHRT
QLESHRFEKRGKCPDPISSTIIGELFLPGISDVYGRGAPVCTEFLRVNRNAAHRT
EPDSHLEKPKFVSGYMDPDNDHDDNKVYLFPEKALAEIETHAIITRYRVCYN
DMGGRIVVAKMSTFLKTRLCVSPGVRNGIDTFLDELEDFLLQTRKNRPVLEGLS
TNSNIRGVACIVYHMAIVRAENGPRNGEYEWALYEGVYPPRPGSCAKVNG
GLYTTTKRDYDEAVHARSHPLMOPKIPVHKRPIIVYTDGKYLQIAVDREAEDG
QYVLFITGNDGIVLKVITITINQETSEMEVLEELQVFPVPIELSMETISRKQOLY
IGTESYIAOVKEFHOCMDYACADCCARPCYAMOSISGRVPTGMOAKRRRRDD
VRRGNAQQCEEQOFIGEVLEKTEBERLYVTEINSTLLETPTLQKVMVYQRAHE
TKREVKYTDRIIKMDGLFLKLRHLDAGTYFOIVESHIVHTVVKITLLEIEEYV
DEMSEKDYEEIISHKPCPMOSNIPQSKPWYKEFIQLIGSNFQREVECEKWCYTD
KKRKKLMSPSKMYAMPORODOKKARIRPEHYRLPRNIADS"

BASE COUNT      802 a      518 c      620 g      675 t
ORIGIN
Alignment Scores:      1,93e-73      Length:      2615
Pred. NO.:      2033.50      Matches:      401
Score:      65.13%      Conservative:      122
Percent Similarity:      49.94%      Mismatches:      251
Best Local Similarity:      42.85%      Indels:      29
Query Match:      5      Gaps:      11
DB:

US-09-813-290-2 (1-875) x AF022947 (1-2615)
QY      84  ARGATGAGLARGCYSPROGLNHPROSERMETALAPROSERALATPRLALIECYSTRP 103
DB      4  AAGAGAGCAAGATTTCATCTGGGAGAGATGCATCGCCAGACCTTTTAATTCG 63
QY      104 LeuLeuGlyGlyLeuLeuLeuHisGlySerSerGlyProSerProGlyProSerVal 123
DB      64 GCACATATGCGCTGTACTAGATTACCGGCTGGATATCACGCAACAGATACGACAG 123
QY      124 ProArgLeuArgLeuSerGlyrArgAspLeuLeuSerAlaAspArgSerAlaIlePheLeu 143
DB      124 CCAGAGTTACGCTGTCACATTAAGAGCTGTGGAGTTTAAACGAGCGTCAGCTTCAT 183
QY      144 GlyProGlyGlySerLeuAsnLeuGlnAlaMetTyrLeuAspGlyLeuTyrArgAspArgLeu 163
DB      184 AGCCCATTTGGGATCTTGGTCTTCATATATGCTCTCGATGAATATCAAGAGCAGCTG 243
QY      164 PheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeuAspGlnAlaIleTyrProAspPro 183
DB      244 TTCGTGGAGAGAGACCTCTTGTACTCCCTCAGTTGGATCGAATCAGCAACACTAT 303
QY      184 ArgGluValLeuTyrProProGlnProGlyGlnArgGluGluCysValAlaGlyGlyArg 203
DB      304 CGAGAGATTCATTGGCTTACACTCTTCMAAGCAGANAATGCAATTAATAAGAGAGA 363
QY      204 AspProLeuThrGluCysAlaAsnPheValArgValLeuGlnProHisAsnArgThrHis 223
DB      364 GAT---GCTGATGAATGCTGTAATATATGCTGCTTACACCGATCAATATAGAACCAT 420
QY      224 LeuLeuAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeuIleThrValGlyHis 243
DB      421 CTCTGCTGCTTGGACAGAGAGCTTGTGATCCACTTGTACGTTTATAAAGGTTGGACAT 480
QY      244 ArgGlyValLeu---HisValLeuHisLeuGlnProGlySerValGluSerGlyArgGlyArg 262
DB      481 CCATCAACAGAGATCATCTGTTTCACTGCAATCGCACAATTTGGAGAGAGAGGAGCGAG 540
QY      263 CysProHisGluProSerArgProPheAlaSerThrPheIleAspGlyGluLeuTyrThr 282
DB      541 TGCCCTTTTGGACCTCTCTCTCTCACTTCCTCACTTAATTTGGTGGTGAATTTTACT 600
QY      283 GlyLeuThrAlaAspPheLeuGlyArgGluAlaMetIlePheArgSerGlyGlyProArg 302
DB      601 GGCTCTACACTGACTACTGGGAGAGATGCTGCTGCTTTCGACACATGAAATCGTATG 660

```

```

QY      303 ProAlaLeuArgSerAspSerAsp---GlnSerLeuLeuHisAspProArgPheValMet 321
DB      661 GCACATCTCCGAAGTGAACCTGATATGACACCTGCGTGAAGAAACGAAAGTGTGTTGT 720
QY      322 AlaAlaArgIleProGlyAsnSerAspGlnAspAsnAspIleValTyrPhePhePheSer 341
DB      721 TCATACATGATCTCTGACAAATGAAAGCACCATGACATTAACAAAGTATCTCTTTTACT 780
QY      342 GluThrValProSerProAspGlyGlySerHisValThrValSerArgValGlyArg 361
DB      781 GAAAAACATTAGAGCTGAG---ACANGCACATCATGCCATTTTACACAGAGTGGAGC 837
QY      362 ValCysValaAsnAspAlaGlyGlyGlnArgValLeuValAsnLysTyrSerThrPheLeu 381
DB      838 GTGTGTGTAATGATATGAGGAGAGACAGCGGATCGTTGGAATTAATGAGACCTTCTC 897
QY      382 LysAlaArgLeuValCysSerValProGlyProGlyGlyAlaGluThrHisPheAspGln 401
DB      898 AAACACAGACTAGTTGTCTGCTGCTGAGAAATGGAAATGGATATGATACCATTTGATGA 957
QY      402 LeuGluAspValPheLeuLeuTyrProLysAlaGlyLysSerLeuGluValTyrAlaLeu 421
DB      958 TTAGAGATGATGTTTGTGCTGCAACTGCTGATTAACAAAATCCAGTGATATTGGCTTC 1017
QY      422 PheSerThrValSerAlaValPheGlnGlyPheAlaValCysValTyrThrHisMetAlaAsp 441
DB      1018 TTCACGACTACACAGAAATATATTAGAGATATGCTATATGTTTACCATATGGCAATC 1077
QY      442 IleTyrGluValPheAsnGlyProPheAlaHisArgAspGlyProGlnHisGlnTyrGly 461
DB      1078 GTCCGAGCAGCTTTCATATGACCATATGCTCATTAAGAAAGACCAANAATACATCTGGCT 1137
QY      462 ProTyrGlyGlyLysValProPheProArgProGlyValCysProSerLysMetThrAla 481
DB      1138 CTGTACGAGAGGAAATGACTATCTAGCCCTGCTGCTCATGTGCCACAAAGTAAT--- 1194
QY      482 GlnProGlyArgProPheGlySerThrLysAspTyrProAspGlnValLeuGlnPheAla 501
DB      1195 -----GGCGCTGTGATACCTACCAACCAAGACTACCTATATAGCTGTGCATTTTGC 1248
QY      502 ArgAlaHisProLeuMetPheTyrProValArgProArgHisGlyArgProValLeuVal 521
DB      1249 AGAGATCATCTACTAATGATATGATACGCCATGACCGCTGTCATTAAGAACCAATTCAGTA 1308
QY      522 LysThrHisLeuAlaGlnGlnLeuHisGlnIleValValAspArgValGluAlaGluAsp 541
DB      1309 AAAACAGATGGAAAGTACAACTTAAGCAATAGCTGTGACAGAGTGAAGCTGAAGAC 1368
QY      542 GlyThrTyrAspValIlePheLeuGlyThrAspSerGlySerValLeuLysValIleAla 561
DB      1369 GGGCAATATATGTTGTTGATTCATGGCAGATATATGGAATGTGCTGGAAGTCAATTACA 1428
QY      562 LeuGlnAlaGlyLysSerAlaGluProGluGluValValLeuGluGluValAlaPhe 581
DB      1429 ATTACATATCAAGACAGACAGAAATCAATGAGAAAGTCTTCTTGAAGAGCTCCAGTATTC 1488
QY      582 LysValProThrProIleThrGlnMetGluIleSerValLysArgGlnMetLeuTyrVal 601
DB      1489 AAGGTGCCAATTCCTATCTTCCATGGAATTCCTTCAAAAATGCAACACCTCTACTAT 1548
QY      602 GlySerArgLeuGlyValAlaGlnLeuArgLeuHisGlnCysGluThrTyrGlyThrAla 621
DB      1549 GGAATGAGTGAAGTATGATGCAAGAGTTCACACAGTGTGATGATATGAGCAGACACC 1608
QY      622 CysAlaGluCysCysLeuAlaArgAspProTyrCysAlaIlePheAspGlyValSecCysThr 641
DB      1609 TGTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
QY      642 HistTyrArgProSer-----LeuGlyLysArgArgPheArgArgGlnAspIleArgHis 659
DB      1669 CGGTACTACCCACAGAGATGACGCAAAAGAGAGCTTCGAGAGCAAGATGTACGACAT 1728
QY      660 GlyAsnProAlaLeuGlnCysLeuGlnSerGlnGluGluAlaValAlaGlyLeuVal 679

```

```

Db 1729 GGAAGAGCAGCTGAGCAGCTGTTGGCAGCAGTTCATGAGAGAGCTGGAGAACT 1788
Oy 680 AAlaAlaTherMetValTyrGlyThrGluHisSerTherPheLeuGlyCysLeuProLys 699
Db 1789 GAGGAGCAGATGGTTTATGAGTAAGAGTACAAATGACCCCTTGGAGTACACCCCTGG 1848
Oy 700 SerProGlnAlaAlaValAlaTyrPheLeuGlnArgProGlyAspGlnGlyProAspGln 719
Db 1849 ACCCTTACAGGCAAAAGTAAAGTGGTTGTTCACAGCAGCCATGAACTAGAGAGAAAG 1908
Oy 720 ValLysTherAspGlnArgValLeuHisThrGluArgGlyLeuPheArgLeuSer 739
Db 1909 GTGAGACAGATGAAAGAAATCAATCAAAATGACCTGCGCTCTTAATTTGAGAGCTGAT 1968
Oy 740 ArgPheAspAlaGlyThrTyrThrCysThrThrLeuGlnHisGlyPheSerGlnThrVal 759
Db 1969 CGGCTGATGACAGGAGCATTATTTTGTGACAGAGCTGAGACACAGCATAGTCACTGTT 2028
Oy 760 ValArgLeuAlaLeuValValIleValAlaSerGlnLeuAspAsnLeuPheProGlu 779
Db 2029 AGAAATATACCTCGAATAGTGTAGAGAAAGCTGTAGTAAATGTTCACTAAAGAC 2088
Oy 780 ProLysProGlu-----GluProAlaArgGlyGlyLeuAlaSerThr 794
Db 2089 TATGAGGAGAGATATCTCAAAATGCCATGCCATGACAGCAACATACCTCAGGTA 2148
Oy 795 ProProLysAlaTyrTyrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuProArg 814
Db 2149 ---TCAAAACCATGTGACAAAGAAATTTCTTCAACTAGTATGATACAGCACTCCAGAG 2205
Oy 815 ValAspGlnTyrCysGlnArgValTyrCysArgGlyThrThiCysSerGlyCysPhe 834
Db 2206 GTGAGAGATACTGTGAAAAAGCTGGTGTACAGAT----- 2241
Oy 835 ArgSerArgSerArgGlyLysGlnAlaArgGlyLysSerThrAlaGlyLeuGlnGly 854
Db 2242 AAGAGAGAGAAAGCAAGCTGAATATGCTCCATCCAG---TGAAATATGCCATCCTCAG 2298
Oy 855 LysLys-----MetLysSerArgValHisAlaGlnHisAsnArgThrPro 869
Db 2299 GAGAAAGAGCAGATCAGAGAGAGAGAGATCAGGACAGCAGCATACCGGTTGCCG 2358
Oy 870 ArgGluVal 872
Db 2359 AGGACATA 2367

RESULT 14
LOCUS AB002329 6474 bp mRNA linear PRI 06-OCT-2001
DEFINITION Human mRNA for KIA0331 gene, complete cds.
ACCESSION AB002329
VERSION AB002329.1 GI:2224602
KEYWORDS KIA0331.
SOURCE Homo sapiens male brain cDNA to mRNA, clone lib:pbluescript SK
plus clone:H60928.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 6474)
Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
VII. The complete sequences of 100 new cDNA clones from brain which
can code for large proteins in vitro
DNA Res. 4 (2), 141-150 (1997)
9734984
2 (bases 1 to 6474)
Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
Direct Submission
Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute,
1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)

```

```

FEATURES
    source
        Location/Qualifiers
            1..6474
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="HG0928"
                /sex="male"
                /tissue_type="brain"
                /clone_lib="pbluescript SK plus"
            467..2794
                /gene="KIA0331"
            467..2794
                /gene="KIA0331"
            /codon_start=1
            /protein_id="BAA20789.1"
            /db_xref="GI:2224603"
            /translation="MASAHITITLLMKYLELWKGHTADTTTPRLRISKEILNLN
            RSTIRSPGELDLTMDLEYOERLFEVGRDLVSLSEIRISDEKELHWSTALMK
            EECIMKGADGECANVRYLHHYNRHLTCTGAFEDPYCAFIRVGYLHEDLPHLES
            PRSEKRGRCDFPSSSFLISLEFAGLYSDYSRDAIFRSMGLAHLIRTEHD
            ERLKEPKFVGSYMIIPDNEDRDNDKVFYFEETKALEAENNAHATYRVGRILCVNG
            ORILVNMSTFLKARLVCSVPMNGIDYFDELEDFVLTLPDHRNPYFGLPNTSN
            IFRGHAICVYHMSSTRAAFNGSPYAKKEGEYHWSYEGKVPIPRGSCSKVNGRYG
            TTKQYDDAIFRASHPLMTQAIKFRHKPVLVKTGDKYNLQIAVDREALDGYDV
            LFTGNDGIVLKVITYNDEMSEEVILLEQIIRKDPVPIISMSISSKQDLYGSA
            SVAQVRFHHCMDYSGACADCCCLARDPDMGISCRRYPGTGAKRRFRDVRHG
            NAAOOCFGQOPVGDALDKTEHLAGIENNSLLCTPSPLOAKYIWFQKRETRKE
            EVKTDREVYKMDGLFLRLHKSDAGTEFCCTVERSPFVHTVKITLVEVEEKVDMF
            NKDDESDRHMRPCPAOSSISOGAPDWKKEPFLQIIGSFQVFEYCEKVMCTDKRK
            KIKMSPEKKVYNPOEKIRSRPRYRLPRHTLDS"

BASE COUNT  2107 a  1193 c  1294 g  1880 t
ORIGIN

Alignment Scores:
Pred. No.:      1,77e-72      Length:      6474
Score:          2018.50      Matches:      396
Percent Similarity: 66.84%      Conservative: 130
Best Local Similarity: 50.32%      Mismatches:  236
Query Match:    42.53%      Indels:      25
DB:              9      Gaps:      10

US-09-813-290-2 (1-875) x AB002329 (1-6474)
Oy 93 SerMetAlaProSerAlaThrPalaIleCysTrpLeuLeuGlyCysLeuLeuHisGly 112
Db 464 AGCATGGCATCCGGGGGACATATATCACCCTGCTCTGGGGTACTTACGAGCTT 523
Oy 113 GlySerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTyrArgAsp 132
Db 524 TGGACAGAGAGTCAATACAGCTGATACCCACCCCGCTTACGCTGCACATAAAGC 583
Oy 133 LeuLeuSerAlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGln 152
Db 584 CTCTTGAAATCTGACAGCAACATCAATATTTCATACGCCCTTTGGATTTCTTATCTCCAT 643
Oy 153 AlaMetTyrLeuAspGlnIuYrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyr 172
Db 644 ACAATGCTGCTGGATGAGAAATATCAAGAGAGAGCTCTTCGTGGAGGACAGGACCTGTATAT 703
Oy 173 SerLeuArgLeuAspGlnAlaIleThrProAspProArgGlyValLeuTyrProGlnPro 192
Db 704 TCCCTCAGCTTGGAGAGATCAGTACGCGCTATAAAGAGATACACTGGCGGAGTACAGCT 763
Oy 193 GlyIleArgGlnGlyCysValArgAspProLeuThrGlnCysAlaAsnPhe 212
Db 764 CTAAATAAGGAAGATGCAATATGAAAGGAAAGAT---GCGGTGAATGTGCCAATATAT 820
Oy 213 ValArgValLeuGlnProHisAsnArgThrHisLeuLeuAlaCysGlyThrGlyAlaPhe 232
Db 821 GTTCGGGTTTTCATCACTATATACAGGACACCTTTCGACCTGTGATACAGGACCTTTT 880
Oy 233 GlnProThrCysAlaLeuIleThrValGly---HisArgGlyGlnHisValLeuHisLeu 251
Db 881 GATCCAGTTTGTGCTTCATCAGAGTGTGATATATTTGGAGAGATCCTGTGTTCACCTG 940

```


QY	252	GLuPProGlySerValGluSerGlyValGArgCysProHisGluProSerArgProPhe	271
Db	941	GAATTCACCCAGATCTGTGAGAGAAAGGGCAGATGCTCTTTGGACCCAGCTCTCTTC	1000
QY	272	AlaSerThrPheIleAspGlyGluLeuTyrThrClyLeuThrAlaAspPheLeuGlyArg	291
Db	1001	ATCTCCACTTAAATGTGTAAGTATTTTGTCTGGACTCTACAGTACTACTGGACGAGA	1060
QY	292	GLuAlaMetIlePheArgSerGlyGlyProArgProAlaLeuArgSerAsp--SerAsp	310
Db	1061	GAGCGTCGATCTCCGACAGATGGGCGACTGGCCCTATCCGACACAGACATGACGAT	1120
QY	311	GlnSerLeuLeuHisAspProArgPheAlaMetAlaAlaArgIleProGluIleAsnSerAsp	330
Db	1121	GAGCGTCGTGGTGAAGAACAATTTGTAGCTTCATCATGATTCCTGTGACATATGAAC	1180
QY	331	GlnAspAsnAspLysValTyrPhePheSerGluThrValProSerProAspGlyGly	350
Db	1181	AGAGATGACACAAAGATATATTTCTTTTACTGGAAGGACCTGGAGGACGAATAACAT	1240
QY	351	SerAsnHisValThrValSerArgValGlyArgValCysValAsnAspAlaGlyGln	370
Db	1241	GCT--CACGCAATTTTACACACAGGTCGGCGGACCTGTGTGAAGATGATAGAGGGCG	1297
QY	371	ArgValIleValAsnLysTyrSerThrPheLeuLysAlaArgLeuValCysSerValPro	390
Db	1298	AGAATACCTGGTGAATTAAGTACGACACTTCTCTAAAGGAGACCTGTTGCTCAGTACA	1357
QY	391	GlyProGlyGlyAlaGluThrHisPheAspGlnLeuGluAspValPheLeuLeuTrpPro	410
Db	1358	GGAATGGAATGGAATGACATATTTTGTATGATTAAGAGAGCGTTTCTTCTACTACC	1417
QY	411	LysAlaGlyLysSerLeuGluValTyrAlaLeuPheSerThrValSerLValAlaPheGln	430
Db	1418	AGAGATCATTAAGAAATCCAGTGATTTTGGACTCTTAACTACACAGATATATTTTGGCA	1477
QY	431	GlyPheAlaValCysValTyrHisMetAlaAspIleTrpGluValPheAsnGlyProPhe	450
Db	1478	GCGCATGCTATATGCTGTCTATACATGTCTACATTCGGGACGCTTCAACGACCATAT	1537
QY	451	AlaHisArgAspGlyProGlnHisGlnTrpGlyProTyrGlyGlyLysValProPhePro	470
Db	1538	GCACATTAAGGAAGACCTCGAATACCACTGGTACGTATGAAGGAATAATCCCTTATCCA	1597
QY	471	ArgProGlyValCysProSerLysMetThrAlaGlnProGlyLysArgProPheGlySerThr	490
Db	1598	AGCGCTGCTTCTGTGGCCAGCAAAAGTAAT-----GGAGGAGATATGGGAACACAC	1648
QY	491	LysAspTyrProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpPro	510
Db	1649	AAGGACTATCTCGATGATGCCATCCGATTTTGCAGAAAGCATCCATCAATGAATGCCAGGCC	1708
QY	511	ValArgProArgHisGlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHis	530
Db	1709	ATTAATACCTGCCCATTAATAAACCAATATTTGTTAAAAACAGATGGAAATATTAACCTGA	1768
QY	531	GlnIleValValAspArgValGluAlaGluAspGlyThrTyrAspValIlePheLeuGly	550
Db	1769	CAATATGCAGTATGATGAGCTGGAGCTGGAGATGGCCAAATATATACGCTCTTTATTTGGG	1828
QY	551	ThrAspSerGlySerValLeuLysValIleAlaLeuGlnAlaGlyLysSerAlaGluPro	570
Db	1829	ACAGATAATGGAATTTGCTGGAAGTAATCATCAATTTTACACCAAGAAATGGAATCATG	1888
QY	571	GluGluValValLeuGluGluLeuGlnValPheLysValProThrProIleThrGluMet	590
Db	1889	GAGAAGTATATCTACAGAAGACTTCAGATTTTCACAGATCCAGTTCTCTATTTATTTATG	1948
QY	591	GluIleSerValLysArgGlnMetLeuTyrValGlySerArgLeuGlyValAlaGlnLeu	610
Db	1949	GAAATTTCTTCAAAAGGCGCAACGCTGTATATTTGATCTCTTCTGCTGGCTCAAGTGC	2008

Qy	611	ArgLeuHisGlnCysGluThrTyrGlyValAlaCysAlaGlnCysGluLeuAlaIleAsp	630
Db	2009	AGATTTCACATCCTGTGCATGTATGAAAGTCCTTGCTGCTACATGCTCCGCTGGCTGAGAC	2068
Qy	631	ProtyrCysAlaIleTyrAspGlyAlaSerCysThrHisTyrArgPro-----SerLeuGly	648
Db	2069	CGTTACTGTGCTGGGATGGCATATACCTGCTCCCGGATATTAACCAAGACACACATGCA	2128
Qy	649	LysArgGlyPheArgArgGlnAspIleArgHisGlnLysProAlaLeuGlnCysLeuGly	668
Db	2129	AAAAGCGCTTCCGAGACAAAGATGTTGTGACATGGAATGACGTCAGAGTGGCTTGGGA	2188
Qy	669	GlnSerGlnGlnGlnAlaValGlyLeuValAlaIleThrMetValIleTyrGlyThrGlu	688
Db	2189	CAACAGTTTGTGGGAGCTTTGGATATAGCATAGCAACATCTTGCTTATGGCATAGAG	2248
Qy	689	HisAsnSerThrPheLeuGlnCysLeuProCysSerProGlnAlaIleValAlaArgTyrLeu	708
Db	2249	AACAACAGTACTTCTGCTGGAATGATCCGACAGTCTTTACAAAGGAAATATATCTGTTT	2308
Qy	709	LeuGlnArgProGlyAspPheGlyProAspGlnValIleTyrAspGluAlaValLeuHis	728
Db	2309	GTACACGAAAGACGCTGACACAAAGAAAGAGGAGGAGGAGAGATGACAGAGTGGTTAAG	2368
Qy	729	ThrGlnArgGlyLeuLeuPheArgTyrGluSerArgPheAspIleGlyThrTyrThrCys	748
Db	2369	ATGGACCTTGGTTTACTCTTCTTAAGGTTTACACAATACAGTGGTGGACCTATTTTGGC	2428
Qy	749	ThrThrLeuGlnHisGlyPheSerGlnThrValValAlaGlnLeuAlaLeuValIleVal	768
Db	2429	CAGACAGTAGACATAGCTTTGTCATACGGTCCGTAATAAACCTTGGAGGTAGTGGA	2488
Qy	769	AlaSerGlnLeuAspAsnLeuPheProProGluProGluProGluGluPro-----	785
Db	2489	GAGGACAAAGTCGAGGATVGTTTAAACAAGGACGTAGAGGAGACAGCATCACAGATG	2548
Qy	786	-----ProAlaArgGlyGlyLeuAlaSerThrProProLysAlaIleTyrTyrLysAspIle	803
Db	2549	CGTTGTCCTGCTCAGATGATGATC---TCGCGAGGAGACAAACATGTGTACAGGAATTC	2605
Qy	804	LeuGlnLeuIleGlyPheAlaAsnLeuProAlaGlyAlaAspGluTyrCysGlnArgValTyr	823
Db	2606	TTTCACCTATCGTTATCCAGAACTTCCAGAAAGTGAAGCAATACTCGGAGAAAGTATGG	2665
Qy	824	CysArgGlyThrThrGluCysSerGlyCysPheArgSerArgSerArgGlyLysGlnAla	843
Db	2666	TGCACAGAT-----AGAAAGACGAAAGAAAGTTAAATGTCA	2701
Qy	844	ArgGlyLysSerTyrPheAlaGlyLeuGlyIleLysMetLysSerArgValHisAla	863
Db	2702	CCCTCCAAAGTGAAGTATGCCAACCTCCAGAGAAAGAAAGCTCGTGTCCAA-----CCT	2755
Qy	864	GluHisAsnArgThrProArg	870
Db	2756	GAGCATTACCGCTCGCCAGG	2776
RESULT 15			
LOCUS	MMZ3948	3988 bp	linear
DEFINITION	M.musculus mRNA for semaphorin Hv (3988 bp).		
ACCESSION	Z93948		
VERSION	Z93948.1		
KEYWORDS	semaphorin H.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1. (bases 1 to 3988)		
AUTHORS	Christensen, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-Apr-1997) Christensen C., Danish Cancer Society,		
	Molecular Cancer Biology, Strandboulevarden 49, Copenhagen,		
	Sealand, Denmark, 2100		

QY 412 AlaGlySerLeuGluValTyrAlaLeuPheSerThrValSerAlaValPheGlnG 431
 DB 1564 GATCTTAAGAAATCCAGTGTATTTGGAGCTGTTTAACTACCGAGCAATATTTAGAGCC 1623.
 QY 432 PheAlaValCysValTyrHisMetAlaAspIleTrrGluValPheAsnGlyProPheAla 451
 DB 1624 CATGCTGATGATGCTATACATGTCAGATATCCGGAAAGCCTTAAATGCGCCATATGCT 1683
 QY 452 HisAlaAspGlyProGlnHisGlnTrrProGlyGlyGlyValProPheProArg 471
 DB 1684 CATAAAGAAGGCCCTGAATACACGCTACTATGAAAGAAAGTCCCTACCAAG 1743
 QY 472 ProGlyValCysProSerIleMetThrAlaGlnProGlyArgProPheGlySerThrIys 491
 DB 1744 CCTGCTTCCTGCGCAGCAAAAGTAAAC-----GAGGCAAGATGAGCAACCA 1794
 QY 492 AspTyrProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetThrProVal 511
 DB 1795 AGATTACCCGATGACGCGATCCGTTCCAGAGATGATCCTATATGATACAGCCATA 1854
 QY 512 ArgProAlaGlnHisGlyArgProValLeuValLysThrHisLeuAlaGlnLeuHisGln 531
 DB 1855 AAACCTGTTCAATAAAACCAATACCTGTAACAAACAGATGAAATACACCTGAGGCA 1914
 QY 532 IleValValAspArgValGluAlaGluAspGlyThrTyrAspValIlePheLeuGlyThr 551
 DB 1915 CTGGCCGATGATGCGGATGCGAGGATGCGCAGATGAGCTGATATTATGTCACA 1974
 QY 552 AspSerGlySerVal---LeuLysValIleAlaLeuGlnAlaGlySerAlaGluPro 570
 DB 1975 GACACAGCAATGCTGCTGTAAGATTAACACATTTACACCAAGAAACAGATGATG 2034
 QY 571 GluGluValValLeuGlnGluLeuGlnValPheLysValProThrProIleThrGluMet 590
 DB 2035 GAGGAAGTATTCTAGAGAACTTCAAAATATTCAGATCCAGCCCTATCATTTCTATG 2094
 QY 591 GluIleSerValLysArgGlnMetLeuTyrValGlySerArgLeuGlyValAlaGlnLeu 630
 DB 2095 GAAATTTCTTCAAGAGCAACAGACGTTTCATTTGATCGCCTGCTGTGGCAGAGTG 2154
 QY 611 ArgLeuHisGlnCysGluThrTyrGlyThrAlaCysAlaGlnCysCysLeuAlaArgAsp 630
 DB 2155 AGATTCCATCATCTGCGCATGATGTCAGATGCTGTGCTGACGTGCGCTGCGTCAGAC 2214
 QY 631 ProTyrCysAlaTrrAspGlyLysSerCysThrHisTyrArgProSerLeuGly----- 648
 DB 2215 CCGTACTGTGCTGGGATGCGATATCTCTGCTCCAGTACTACCCAAAGGTGCACAGAA 2274
 QY 649 LysArgArgPheArgArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCysLeuGly 668
 DB 2275 AAGAGAGAGTCCGCGAGGAGGAGCGTTGGCATGCGCAACGCCCAACAGATGCTTGG 2334
 QY 669 GlnSerGlnGluGluAlaValGlyLeuValAlaAlaThrMetValTyrGlyThrGlu 688
 DB 2335 CAGCAATTTGTTGGAGCCGTTGGCAGAGCTGAGAGAGAGAGCTGATGCGATAGAG 2394
 QY 689 HisAsnSerThrPheLeuGlnCysLeuProLysSerProGlnAlaAlaValArgTrrPleu 708
 DB 2395 AGCAACAGTACTGTTGGATGACACCCGCTATCATCAACAGCAAAAGCATCTGCTTT 2454
 QY 709 LeuGlnArgProGlyAspGluGlyProAspGlnValLysThrAspGluArgValLeuHis 728
 DB 2455 CTACAGAAAGGAGCGAGCGTAAGAAAGAGAGGTAGAGCGATGACAGAGTTGTCAAG 2514
 QY 729 ThrGluArgGlyLeuLeuPheArgGluArgLeuSerArgPheAspAlaGlyThrTyrThrCys 748
 DB 2515 AAGGACTTGGGCTTGTCTTCCACAGATACGCAAGTACAGATGACAGGACCTATTTTGC 2574
 QY 749 ThrThrLeuGlnHisGlyPheSerGlnThrValValArgLeuAlaLeuValIleVal 768
 DB 2575 CAGACAGTACACAAATTTTGTCCATCTGTGCTGTAATAATCACTTGGAGGTGTCGAA 2634

QY 769 AlaSerGlnLeuAspAsnLeuPheProProGluProLysProGluGlu----- 784
 DB 2635 GAGCATTAAGTGTGAGGCGCATGTTTCAATAGACCATGAGAGGAAAGACATCACAAGATG 2694
 QY 785 -----ProProAlaArgGlyGlyLeuAlaSerThrProProLysAlaTrrPylLysAsp 802
 DB 2695 CCGTGCCCTCCCTTAAGCCGATGCTGTCAGGGGACAAACCG-----TGTATCAAGAGAA 2748
 QY 803 IleLeuGlnLeuIleGlyPheAlaAsn---LeuProArgValAspGluThrCysGluArg 821
 DB 2749 TTTCTTGCACTATTGCTACACAGCAGCAGAGTTCAGAGAGTGAAGATATCTCGCAAAAG 2808
 QY 822 ValTrrCysArgGlyThrThrGluCysSerGlyCysPheArgSerArgSerArgGlyLys 841
 DB 2809 GTGTGATACAGAT-----AAGAAAGCAAAAGCTTAA 2844
 QY 842 GlnAlaArgGlyLysSerTrrAlaGlyLeuGlnLeuGlyLysLysMetLysSerArgVal 861
 DB 2845 AAGTCTCCCTCCAAAGTGAAGATGCGCAACCCCAAGAAAGAGGCTGCTTAA 2901
 QY 862 HisAlaGluHisAsnArgThrProArg 870
 DB 2902 ----GCTGACACTTCGCGCTGCCAGG 2925

Search completed: October 9, 2002, 21:11:08
 Job time : 3203 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 9, 2002, 18:51:59 ; Search time 242 Seconds
(without alignments)
6207.852 Million cell updates/sec

Title:	US-09-813-290-2
Perfect score:	4746
Sequence:	1 MACALAGKVPFPGSMRYVHK.....KKMSRVHAENRTPREYEAT 875

Scoring cable:		
BLOSUM62		
Xgapop	10.0 ,	Xgapext 0.5
Ygapop	10.0 ,	Ygapext 0.5
Fgapop	6.0 ,	Fgapext 7.0
Delop	6.0 ,	Delext 7.0

```
Searched:      1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

Command line parameters: -DEV-xlh
-MODE=frame-pzn,model
-o/cen2.1/USPTO_spool/US098313290/runat.09102002.094645.19112/abp_query.fasta.1.1033
-DB=N_Geneseq.032802 -OFMT=fastap -SUFFIX=eng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -INITS=b1ts -START=1 -END=-1 -MATRY=b1os1000 -TRANS=human4.0 -CB
-LIST4=5 -DCCALIGN=200 -THR_SCORE=pt -THR_MAX=100 -THR_MIN=0 -ALLEN=15
-MODE=LOCAL -CONFMT=pt -NORM=ext -HEAPSIZED=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US098313290.ecgn.1.1.74.urnat.09102002.094645.19112 -NCP=6 -ICPD=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEDOUT=120
-WARN_TIMEDOUT=30 -THREADS=1 -XGAPOP=0.5 -XGAPEXT=0.5 -XGAPOP=6 -XGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
Database :
1: N.Geneseq_032802.*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	4746	100.0	2628	AAH47790	Novel human protel
2	4746	100.0	3568	AAH47792	Novel human protei
3	4357	91.8	4859	AAA93109	Human secreted pro
4	4215	88.8	2349	AAAH7791	Novel human protei
5	4046.5	85.3	2340	AAP00250	Nucleotide sequenc
6	3102.5	65.4	2337	AAFP0251	Nucleotide sequenc
7	2056	43.3	2898	AAV55367	Human semaphorin E
8	1955	41.2	3988	AAZ28469	Mouse semaphorin H
9	1834	38.6	2709	AAH47049	Semaphorin D CDNA
10	1830	38.6	2530	AAAT5767	Human semaphorin I
11	1830	38.6	2601	AAO6B442	Human semaphorin I
12	1721	36.3	2331	AAAG8112	Human brain tissue
13	1721	36.3	3871	AAAG37109	Human brain tissue
14	1721	36.3	3871	AAAS4098	Human PRO1491 (UNC
15	1721	36.3	3871	AAAF54421	Human DNA encoding
16	1721	36.3	3880	AAAX89113	DNA encoding prote
17	1721	36.3	2975	AAAX03792	Human brain tissue
18	1538	32.4	2975	AAAC66800	Human semaphorin E
19	1538	32.4	1481	AAAO92331	Clone BR33-4 codi
20	1295	27.8	1481	AAAK81569	Human collagen/cd
21	988	20.8	3203	AAAK81569	Human immune/haema
22	983	20.7	3199	AAAC69627	Human semaphorin
23	974	20.5	1444	AAAC69627	Human semaphorin e
24	892	18.8	4391	AAAF77462	Human CD100 DNA se
25	890	18.8	2769	AAAF77462	Murine M-Sema-F CD
26	889	18.8	2769	AAAC92383	Human DNA encoding
27	887	18.7	2769	AAAD08048	Human DNA encoding
28	887	18.7	2769	AAAV31121	Primer #75 used in
29	874	18.4	4157	AAAF77463	Human semaphorin G
30	874	18.4	4157	AAAF77463	Human semaphorin G
31	874	18.4	4157	AAAF77463	Human semaphorin G
32	873	18.4	3503	AAAC92384	Human CD100 DNA se
33	869.5	18.3	3781	AAAF37092	Human CD100 nucleo
34	869.5	18.3	3781	AAAF37092	Murine M-Sema-F CD
35	869.5	18.3	3781	AAAF37092	Human DNA encoding
36	867.5	18.3	2532	AAAB04075	Human semaphorin I
37	867.5	18.3	2613	AAAB04075	Human semaphorin I
38	864.5	18.2	2517	AAAB04077	Human semaphorin G
39	861	18.1	2517	AAAB04076	Human semaphorin G
40	861	18.1	2598	ABA04078	Human semaphorin G
41	857	18.1	2155	AAAC84887	Human SEC6 nucleic
42	857	18.1	2155	AAAC84887	Human SEC6 nucleic
43	857	18.1	2284	AAAC84888	Human SEC7 nucleic
44	857	18.1	2703	AAAF7437	Human SEC7 nucleic
45	855	18.0	3556	AAAD08263	Sequence encoding Human secreted pro

/product= "NHP"

W0200170806-A2.

27-SEP-2001.

20-MAR-2001; 2001WO-US08834.

20-MAR-2000; 2000US-190638P.

22-MAR-2000; 2000US-191188P.

31-MAR-2000; 2000US-193639P.

(LEXI-) LEXICON GENETICS INC.

Walke DW, Milgowski NL, Turner CA, Hilbun E, Wang X, Donoho G, Scoville J.

WPI; 2001-611483/70.

P-PSDB; AAG65619.

New polynucleotides encoding human proteins that share structural similarity with semaphorin proteins, protein hormones of neurohypophyseal family for drug screening, diagnosis and therapy of biological disorders

Claim 1; Page 34-35; 43pp; English.

The invention relates to novel human secreted proteins (NHP) that share structural similarity with semaphorin proteins, protein/peptide hormones of the neurohypophyseal family and oxytocin (neurophysin 1 precursor) family. The NHP nucleotide sequences are useful in drug screening of techniques for treating symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Nucleotide constructs encoding NHP products are useful in gene therapy for modulating NHP expression. The constructs can be used to genetically engineer host cells to express NHP products in vivo, these genetically engineered cells function as bioreactors in the body delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. The sequences also find use in molecular mutagenesis/evolution of proteins that are partially encoded by the NHP sequences. The encoded NHP polypeptides are useful for generating antibodies, as reagents in diagnostic assays, for identifying other cellular gene products related to NHP and as reagents in assays for screening for compounds that are useful in the treatment of mental, biological or medical disorders and diseases. The present sequence represents a cDNA encoding a NHP.

Sequence 2628 BP; 459 A; 836 C; 848 G; 484 T; 1 other;

Alignment Scores:

Pred. No.:	7	19e-252	Length:	2628
Score:	4746.00		Matches:	875
Percent Similarity:	100.00%		Conservative:	0
Best Local Similarity:	100.00%		Mismatches:	0
Query Match:	100.00%		Indels:	0
	22		Gaps:	0

US-09-813-290-2 (1-875) x AHA47790 (1-2628)

QY 1 MetAlaCysAlaLeuAlaGlyValPheProMetGlySerTrpProValTrpHisLys²⁰Db 1 ATGGCTGTGCTCCCTACTGGGAGAGCTTCCCAATGGGAGCTGGCCAGCTGGCACA⁶⁰QY 21 SerLeuHisTrpAlaAsnLysValGluGluAlaAlaGlyLysArgGlnGlyProSer⁴⁰Db 61 AGCTGTGACATGGGCGCAAGAGTGGAGAGGAGGAGGAGTGGAGGAGGAGGAGG¹⁸⁰QY 41 LeuLeuLeuSerSerAlaProLeuProAlaGlnAspTrpValGluProLeuProTyrLys⁶⁰Db 121 CTTA¹⁸⁰QY 61 TrpTrpProGluGlySerTrpAlaAsnLysValArgArgArgProAlaGlyProGluGly⁸⁰Db 181 TGGTGGCTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG²⁴⁰QY 81 SerAlaGlyArgArgGlnArgCysProGlnPheProSerMetAlaProSerAlaTrpAla¹⁰⁰Db 241 TCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG³⁰⁰QY 101 IleCysTrpLeuLeuGlyGlyLeuLeuHisGlyGlySerSerGlyProSerProGly¹²⁰Db 301 ATTTGGCTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG³⁶⁰QY 121 ProSerValProArgLeuArgLeuSerTrpArgAspLeuLeuSerAlaAsnArgSerAla¹⁴⁰Db 361 CCCAGTGTGCCCCGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT⁴²⁰QY 141 IlePheLeuGlyProGluGlnGlySerLeuAsnLeuGlnAlaMetTrpLeuAspGlyTyrArg¹⁶⁰Db 421 ATCTTCTGGGCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁴⁸⁰QY 161 AspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeuAspGlnAlaTrp¹⁸⁰Db 481 GACGCTCTTCTGTGGGTGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁵⁴⁰QY 181 ProAspProArgGluValLeuTrpProGlnProGlnProGlnProGlnProGlnProGln²⁰⁰Db 541 CCAGATCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁶⁰⁰QY 201 LysGlyArgAspProLeuThrGlyCysAlaAsnPheValArgValLeuGlnProHisAsn²²⁰Db 601 AAGGAGAGATCTTGTGACAGAGTGGCCAACTTCTGCGGAGTGTACAGCTCACAC⁶⁶⁰QY 221 ArgTrpHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeuLeuTrp²⁴⁰Db 661 CGAGCCACCTCTAGCTGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁷⁰⁰QY 241 ValGlyHisArgGlyGluHisValLeuHisLeuGluProGlySerValGluSerGlyArg²⁶⁰Db 721 GTTGGCCACCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁷⁸⁰QY 261 GlyArgCysProHisGluProSerArgProPheAlaSerTrpPheIleAspGlyGluLeu²⁸⁰Db 781 GGGCGTCCCTCTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG⁸⁴⁰QY 281 TyrTrpGlyLeuThrAlaAspPheLeuGlyValGluAlaMetIlePheArgSerGly³⁰⁰Db 841 TACAGGCTCTACCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT⁹⁰⁰QY 301 ProArgProAlaLeuArgSerAspSerAspGlnSerLeuLeuHisAspProArgPheVal³²⁰Db 901 CCTGGCCAGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTG⁹⁶⁰QY 321 MetAlaAlaArgIleProGluAsnSerAspGlnAspAsnAspLysValTyrPhePhePhe³⁴⁰Db 961 ATGGCCCGCCGAGTCCCTGAGACCTGACGAGACGAGACGAGACGAGACGAGACG¹⁰²⁰QY 341 SerGluThrValProSerProAspGlyGlySerAsnHisValTrpValSerArgValGly³⁶⁰Db 1021 TCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG¹⁰⁸⁰QY 361 ArgValLysValAsnAspAlaGlyGlyGlnArgValLeuValAsnLysTrpSerTrpPhe³⁸⁰Db 1081 CGGTCTGCTGATGATGATGCTGGGGGCGACGGGCTGTGGAGCAAAATGGACACTTTC¹¹⁴⁰QY 381 LeuLysAlaArgLeuValCysSerValProGlyProGlyGlyAlaGluTrpHisPheAsp⁴⁰⁰Db 1141 CTGAGGCGAGGCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT¹²⁰⁰QY 401 GlnLeuGluAspValPheLeuLeuTrpProLysAlaGlyLysSerLeuGluValTyrAla⁴²⁰Db 1201 CACCTGAGAGATGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT¹²⁶⁰QY 421 LeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCysValTyrHisMetAla⁴⁴⁰Db 1261 CTGTTGAGACCTCAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT¹³²⁰

Sequence 3568 BP: 633 A; 1099 C; 1156 G; 680 T; 0 other;

Alignment Scores:

9,96e-252 Length: 3568
 Score: 476.00 Matches: 875
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-813-290-2 (1-875) x AAH47792 (1-3568)

QY 1 MetaAlaCysValLeuAlaGlyValAlaPheProMetGlySerTrpProValITrphHisLys 20
 Db 555 ATGGCTGTGGCTTACCTGGGAAGTCTTCCCAATGGGGAGCTGGCCAGTGTGGCAAAA 614
 QY 21 SerLeuHisITrpaLAsnLysValGluGluAlaAlaGlyLysArgGlnGlyProSer 40
 Db 615 AGCTTCACATGGGCAACAAAGTGGAAAGAAAGGCGGAGGTGGAGCGCAAGGCCCCAGC 674
 QY 41 LeuLeuLeuSerSerAlaProLeuProAlaGlnAspTrpValGluProLeuProTyrLys 60
 Db 675 CTCCTCTCTCTCTCCGCCCTCTCCGCCAGAGACTGGGTGGAGCCACTGCCTTTAAG 734
 QY 61 TrpTrpProGlyGlySerArgAlaAsnTrpAsnArgArgProAlaGlyProGluGlyGly 80
 Db 735 TGTGGCGCTGTGGGAGCAGCAGCAAACTTCAACCGGCGGCGGAGGAGCAAGAGGGCGGC 794
 QY 81 SerAlaGlyArgArgGlnArgCysProGlnPheProSerMetAlaProSerAlaITrpaL 100
 Db 795 TCTGAGGAGGCGGAGCGGCTGCTCAGTTCCCAAGATGGCCCTCGGCGCTGGGCC 854
 QY 101 IleCysTrpLeuLeuGlyGlyLeuLeuLeuHisGlyGlySerSerGlyProSerProGly 120
 Db 855 ATTTGCTGGCTGTGGGGGCTCTCTCTCCATGGGGGTAGCTGTGGGCCCGCCCGGCCGC 914
 QY 121 ProSerValProAlaGlyLeuArgLeuSerTyrArgAspLeuLeuSerAlaAsnArgSerAla 140
 Db 915 CCCAGTGTGGCGGCTGGGCTCTCTCCATCCGAGACTCTGTGTGCCAACCGCTCTGCC 974
 QY 141 IlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTyrLeuAspGluTyrArg 160
 Db 975 ATCTTGTGGGCGCCAGAGGCTCTCTCAACTCCAGGCGCATGTACTAGTAGAGACCGA 1034
 QY 161 AspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeuAspGlnAlaITr 180
 Db 1035 GACCGCTCTTCTGGGTGGCTGGAGCGCTCTACTCTCTGGGGCTGGAGACGAGCATGG 1094
 QY 181 ProAspProArgGluValLeuTrpProGlnProGlnProGlnArgGluGluCysValArg 200
 Db 1095 CCAATTCCTGGGAGGCTCTGTGGCCACCGGAGGAGAGAGAGGAGAGTGTGTTCGA 1154
 QY 201 LysGlyArgAspProLeuTrpGluCysAlaAsnPheValArgValLeuGlnProHisAsn 220
 Db 1155 AAGGGAAGAGATCTTTGACAGAGTGGCCCACTTGTGGGGGTGTACAGCTCACAAAC 1214
 QY 221 ArgTrpHisLeuLeuAlaCysGlyTrpGlyAlaPheGlnProTrpCysAlaLeuLeuTr 240
 Db 1215 CGAGACCCACTGTAGCTGTGGCTGTGGAGGCGCTTCCAGGCCACTGTGGCCCTCATACA 1274
 QY 241 ValGlyHisArgGlyGluHisValLeuHisLeuGluProGlySerValGluSerGlyArg 260
 Db 1275 GTTGGGCGACCGTGGGAGCATGTGTCTCCACTGTGGAGCTGTGGAGTGGAAATGGCCGG 1334
 QY 261 GlyArgCysProHisGluProSerArgProPheAlaSerTrpPheLeuAspGlyGluLeu 280
 Db 1335 GGGCGGTGGCTTACAGAGCCAGCGCTCTTGGCCAGACCTTCAATAGCGGGAGCGT 1394
 QY 281 TyrTrpGlyLeuTrpAlaAspPheLeuGlyArgGluAlaMetLeuPheArgSerGlyGly 300
 Db 1395 TACAGGGGTGTACAGTGTCTCTGGGGCGAGAGCCATGATCTTCCGAGTGGAGGCT 1454
 QY *301 ProAlaGluAlaLeuArgSerAspSerAspGlnSerLeuHisAspProArgPheVal 320

Db 1455 CCGCGGCGAGCTGTGGCTTCCGACTGTGACGAGTCTTGGCACAGAGCCCGGCTGTGTG 1514
 QY 321 MetaAlaAlaArgIleProGlnAsnSerAlaAspGlnAspAsnLysValITrPhePhe 340
 Db 1515 ATGGCCCCCGGAGTCCCTGAGAACTGTGACGAGACATGACAAAGTGTACTTTCTTTC 1574
 QY 341 SerGluTrpValProSerProAspGlyGlySerAsnHisValITrValSerArgValGly 360
 Db 1575 TCGAGAGAGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1634
 QY 361 ArgValCysValAsnAspAlaGlyGlyGlnArgValLeuValAsnLysTrpSerTrpPhe 380
 Db 1635 CGCTGTGGCTGTATGATGTCTGGGGGCGCACCGGTGTGTGTGAACAAATGAGACACTTC 1694
 QY 381 LeuLysAlaArgLeuValCysSerValProGlyLysValAlaGluTrpHisPheAsp 400
 Db 1695 CTCAGGCGCAGGCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1754
 QY 401 GlnLeuGluAspValPheLeuLeuTrpProLysAlaGlyLysSerLeuGluValITrpaL 420
 Db 1755 CAGCTAGAGATGTGTCTGT 1814
 QY 421 LeuPheSerTrpValSerAlaValAlaPheGlnGlyPheAlaValCysValITrHisMetAla 440
 Db 1815 CTGTTCAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1874
 QY 441 AspIleTrpGluValPheAsnGlyProPheAlaHisArgAspGlyProGlnHisGlnITr 460
 Db 1875 GACATCTGGAGAGTTTTCACAGGGCCCTTGGCCACCGAGATGGGCTTACGACCAAGTAC 1934
 QY 461 GlyProTyrGlyGlyLysValAlaProPheProAlaGlyValCysProSerLysMetTr 480
 Db 1935 GGGCCCTATGGGGGAGAGGTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1994
 QY 481 AlaGlnProGlyArgProPheGlySerTrpLysAspTyrProAspGluValLeuGlnPhe 500
 Db 1995 GCACAGCAGGAGCGCTTGTGGCAGCAGCAAGAGACTACCCAGATGAGGTGTGCGATT 2054
 QY 501 AlaArgAlaHisProLeuMetPheTrpProValAlaProArgHisGlyArgProValLeu 520
 Db 2055 GCCCGAGCCCAACCCCTATGTCTGTGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGG 2114
 QY 521 ValLysTrpHisLeuAlaGlnGlnLeuHisGlnLeuValValAspArgValGluAlaGlu 540
 Db 2115 GTCAAGACCCACTGTGGCCAGCAGCTACCAAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2174
 QY 541 AspGlyTrpTyrAspValIlePheLeuGlyTrpAspSerGlySerValLeuLysValIle 560
 Db 2175 GATGGAGCTTACGATGTCAATTTCTTGGGACTGTACAGAGGTGTGTGTGTGTGTGTGTGT 2234
 QY 561 AlaLeuGlnAlaGlySerAlaGluProGluGluValValLeuGluGluLeuGlnVal 580
 Db 2235 GCTGTCCAGCAGAGGGCTGACGTGAGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2294
 QY 581 PheLysValProTrpProIleTrpGluMetGluLysSerValLysArgGlnMetLeuTr 600
 Db 2295 TTTAAGTGTCCCAACCTTATACCGGAATGAGATCTGTGCAAAAGGCAATGCTATTAC 2354
 QY 601 ValGlySerArgLeuGlyValAlaGlnLeuArgLeuHisLysGluGluTrpTyrGlyTr 620
 Db 2355 GTGGGGCTGTGGGT 2414
 QY 621 AlaCysAlaGluCysCysLeuAlaArgAspProTyrCysAlaITrpaAspGlyAlaSerCys 640
 Db 2415 GCTGTGCAAGAGT 2474
 QY 641 TrpHisTyrArgProSerLeuGlyLysArgArgPheArgArgGlnAspIleArgHisGly 660
 Db 2475 ACCCACTACCGCCCAAGCTTGTGGCAAGCGCGTTCGCGGAGGAGCATCCGGGCGAGGC 2534
 QY 661 AsnProAlaLeuGlnCysLeuGlnSerGlnGlnGluGluAlaValGlyLeuValAla 680

XX	(ALPH-) ALPHAGENE INC.	
PI	Valenzuela D, Yuan O, Hoffman H,	Hall J, Rapiejko P;
XX	WPI, 2000-549267/50.	
DR	P-PSDB; AAB23609.	
XX	New secreted proteins and polynucleotides encoding them, which are	
PT	derived from Homo sapiens, useful for therapy, diagnosis, and research,	
PT	as well as nutritional sources or supplements.	
XX	Claim 26; Page 249-250; 309pp; English.	
PS		
XX	The present sequence is the coding sequence for a human secreted protein.	
CC	The sequence was isolated from an adult brain cDNA library. The proteins	
CC	and coding sequences of the invention can be used in the isolation of <i>in vivo</i> ,	
CC	similar genes and proteins, in the elucidation of their function in have	
CC	and to treat a number of conditions. It is possible that they may have	
CC	uses as nutritional supplements, as cytokine or cell proliferation	
CC	factors. In immune modulation, where they may be used to treat immune	
CC	and autoimmune diseases, as haematopoiesis regulators (treating myeloid or	
CC	lymphoid cell deficiencies), in the promotion of tissue growth, they may	
CC	have chemokine or chemotactic activity, haemostatic or thrombolytic	
CC	activity, or anti-inflammatory activity.	
XX		
SQ	Sequence 4859 BP; 1082 A; 1325 C; 1493 G; 958 T; 0 other;	
	Alignment Scores:	
	Pred. No.: 3.16e-230	Length: 4859
	Score: 4357.00	Matches: 807
	Percent Similarity: 100.00%	Conservative: 0
	Best Local Similarity: 100.00%	Mismatches: 0
	Query Match: 91.80%	Indels: 0
DB:	21	Gaps: 0
US-09-813-290-2 (1-875) x AAA93109 (1'4859)		
QY	69 AsnTYrAsnArgArGProAlaGlyProGlnGlySerAlaGlyArgGlnArgCys	88
DB	4 AACTACAAACGGGGCCGACGAGCAAGGGGGCTCTGCAGCGACGCGACGGCTGC	63
QY	89 ProGlnPheProSerMetAlaProSerAlaTrpAlaIleCysTrpLeuGlnGlyLeu	108
DB	64 CCTCATGTTCCCCACACATGGCCCTCCGGCTCGGGCCATTTGCTGGCTGTAGGGGGCTC	123
QY	109 LeuLeuHISGlyGlySerSerGlyProSerProGlyProSerValProArgLeuArgLeu	128
DB	124 CTGCTCCACGGGGGAGTCTGTGGCCCAAGCCCGGCCCAAGTGGCCCGCTCGGGGCTC	183
QY	129 SerTYrArgAspLeuLeuSerAlaAsnArgSerAlaIlePheLeuGlyProGlnGlySer	148
DB	184 TCTTACCCGAGACCTCTGCTGTGCCAACGCTCTGCATCTTTCTGGGCCCCACAGGGCTCC	243
QY	149 LeuAsnLeuGlnAlaMetTYrLeuAspGluTYrArgAspArgLeuPheLeuGlnGlyLeu	168
DB	244 CTGAACCTCCAGGCGCATGTACTAGATGAGTACCGAGACGCGCTCTTCTGGGTGGCTTG	303
QY	169 AspAlaLeuTYrSerLeuArgLeuAspGlnAlaIleTrpProAspProArgGluValLeuTrp	188
DB	304 GACGGCTCTACTCTCTGGCTGACACGAGGATGCCAGATCCCGGGAGGTCTGTGG	363
QY	189 ProProGlnProGlyGlnArgGlnGlyCysValArgIleGlyGlyTYrArgAspProLeuTrpGlu	208
DB	364 CCACCGCACCCAGGACGAGAGGGAGGAGCTGTGTCGAAGAAGGAGAGATCTTTGACAGAG	423

QY 209 CysAlaAsnProValArgValLeuGlnProHisAsnArgThrHisLeuLeuAlaCysGly 228
 DB 424 TGGCCAACTTGGTGGGGGTCTACAGCTCACAACGGGACCACTGGTAGCTGGCC 483
 QY 229 ThrGlyAlaPheGlnProThrCysAlaLeuLeuThrValGlyHisArgGlyGlnHisVal 248
 DB 484 ACTGGGGCTTCCAGGCCCACTGGCTGCATCAGATTGGCAGCCGGTGGGAGCATGTC 543
 QY 249 LeuHisLeuGlnProGlySerValGlnSerGlyArgGlyArgGlyProHisGlnProSer 268
 DB 544 CTCACCTGGAGCTGGAGCTGGAGAAAGTGGGGGGGGGGCTGCACAGAGCCAGC 603
 QY 269 ArgProPheAlaSerThrPheLeuAspGlyLeuLeuThrGlyLeuThrAlaAspPhe 288
 DB 604 CCTCCCTTGGCAGCCTTCATAGACGGGAGCTGTACAGGGGTCTACTGTGCTTC 663
 QY 289 LeuGlyArgGlyAlaMetLeuPheArgSerGlyGlyProArgProAlaLeuArgSerAsp 308
 DB 664 CTGGGGCGAGAGCCCATGATCTCCAGAGTGGAGGCTCTGGCAGACCTCTGGCTCCGAC 723
 QY 309 SerAspGlnSerLeuLeuHisAspProArgPheValMetAlaAlaArgTLeuProGlnAsn 328
 DB 724 TCTGACCAAGAGCTCTGTGACAGCACCCTGGTTGTGATGGCCCGGATCCCTGAGAAC 783
 QY 329 SerAspGlnAspAsnAspLysValTyrPhePhePheSerGlyThrValProSerProAsp 348
 DB 784 TCTGACCAAGAGCATGAGCAAGGTGACTTCTCTCTGCGAGACGGTCCCTCGCCCAT 843
 QY 349 GlyGlySerAsnHisValThrValSerArgValGlyArgValCysValAsnAspAlaGly 368
 DB 844 GGTGGCTGAAACATGCTACGTGTCAGCCGGCTGGCCCTGCTGGTGAATGATCTGGG 903
 QY 369 GlyGlnArgValLeuValAsnLysTrpSerThrPheLeuLysAlaArgLeuValCysSer 388
 DB 904 GCGCAGGGGGTGGTGAACAAATGAGCACTTCTCCAGGCCAGCTGGTGTCTGCTG 963
 QY 389 ValProGlyProGlyGlyAlaGlyThrHisPheAspGlnLeuGlnAspValPheLeuLeu 408
 DB 964 GTGCCCGCCCTGGTGGTGGCGGAGACCCACTTTGACACAGCTAGAGGATGTCTCTGCTG 1023
 QY 409 TrpProLysAlaGlyLysSerLeuGlnValTyrAlaLeuPheSerThrValSerAlaVal 428
 DB 1024 TGGCCCAAGCCGGGAGAGCTGAGGTGACCGCTGTTCAGACCGTCAAGTGGCTG 1083
 QY 429 PheGlnGlyPheAlaValCysValTyrHisMetAlaAspTLeuProGlyValPheAsnGly 448
 DB 1084 TTCACGGGCTTCCGCTGTGTGTACACATGGCAGACATCTGGAGTTTTCACGGG 1143
 QY 449 ProPheAlaHisArgAspGlyProGlnHisGlnTrpGlyProTyrGlyGlyValPro 468
 DB 1144 CCTTTGCCCAAGAGATGGGCTCAGCACACAGTGGGGCCCTATGGGGGCAAGGTCCC 1203
 QY 469 PheProArgProGlyValCysProSerLysMetThrAlaGlnProGlyArgProPheGly 488
 DB 1204 TTCCCTTGGCTGGCTGGCTGCCAGAGATGACCCGACAGCAGGAGCGCTTTGGC 1263
 QY 489 SerThrLysAspTyrProAspGlnValLeuGlnPheAlaArgAlaHisProLeuMetPhe 508
 DB 1264 AGCACCAGAGACTACCAAGATGAGTGTCTCAATTTGCCCGAGCCCACTCCATGATTTTC 1323
 QY 509 TrpProValArgProArgHisGlyArgProValLeuValLysThrHisLeuAlaGln 528
 DB 1324 TGGCCGTGGCTGAGATGGCCCTGTCTGTCAAGACCACTGGCCGACGAG 1383
 QY 529 LeuHisGlnLeuValAlaAspArgValGlyValGlyAspGlyThrTyrAspValIlePhe 548
 DB 1384 CTACACCAAGATGTGTGTGACCGCTGGAGGAGAGATGGAGCATGATGATTTTC 1443
 QY 549 LeuGlyThrAspSerGlySerValLeuLysValIleAlaLeuGlnAlaGlyLysSerAla 568
 DB 1444 CTGGGGACTGACTCAGGGTGTGTGCTCAAAATCATGCTCTCCAGGACAGGGGCTCAGT 1503

QY 569 GlnProGlnGlnValValLeuGlnGlnLeuGlnValPheLysValProThrProIleThr 588
 DB 1504 GAACCTGAGAGATGCTTCTGAGAGACTCCAGGTGTTTAGGGTCAACACCTATCACC 1563
 QY 589 GlnMetLulleSerValLysArgGlnMetLeuTyrValGlySerArgLeuGlyAla 608
 DB 1564 GAATGAGATCTCTGCAAAAGCAAAATCTATAGTGGGCTGTGGCTGGGTGGCC 1623
 QY 609 GlnLeuArgLeuHisGlnCysGlnThrTyrGlyThrAlaCysAlaGlyCysLeuAla 628
 DB 1624 CAGCTGGGCTGGACCAATGTAGACTTACGGCATGCTGTGCACAGTCTCTGGCC 1683
 QY 629 ArgAspProTyrCysAlaTrpAspGlyAlaSerCysThrHisTyrArgProSerLeuGly 648
 DB 1684 CCGGACCACTACTGCTCTGGAGTGGCTGCTCTCTTACCCACATCCGCCAGCTTGGC 1743
 QY 649 LysArgArgPheArgArgGlnAspIleAlaGlnHisGlyAsnProAlaLeuGlnCysLeuGly 668
 DB 1744 AACGCCGCTTCCGCGGAGAGCATCCGGACAGCCACCTGCTGACGGCACGGAG 1803
 QY 669 GlnSerGlnGlnGlnValAlaValGlyLeuValAlaAlaThrMetValTyrGlyThrGln 688
 DB 1804 CAGAGCCAGAAAGAGAGCAGTGGGACTTGTGGACGCCACCATGCTTACGGCACGGAG 1863
 QY 689 HisAsnSerThrPheLeuGlnCysLeuProLysSerProGlnAlaAlaValArgTrpLeu 708
 DB 1864 CACATAGCATCTCTCTGAGTGGCTGCGCAAGTCTCCACAGCTGCTGTGGCTGGCTC 1923
 QY 709 LeuGlnArgProGlyLysAspGlyLysProAspGlnValLysThrAspGlnArgValLeuHis 728
 DB 1924 TTGCAGAGCCAGGGGATGAGGGGCTGACACAGTGTGAAGACGAGAGGATCTTGGAC 1983
 QY 729 ThrGlnArgGlyLeuLeuPheArgArgLeuSerArgPheAspAlaGlyThrTyrThrCys 748
 DB 1984 ACGGAGCGGGGCTCTGTCCGAGCTTACCGCTTGATGGAGCGGACCTACACCTGC 2043
 QY 749 ThrThrLeuGlnHisGlyPheSerGlnThrValValArgLeuAlaLeuValIleVal 768
 DB 2044 AACACTGTGAGCATGGCTTCTCCAGACTGTGGTCTGCTGGCTGTGTGATTTGTG 2103
 QY 769 AlaSerGlnLeuAspAsnLeuPheProGlnProLysProGlnGlnProProAlaArg 788
 DB 2104 GCGTCACAGCTGGACAACTTCTCCCTCGGAGCAAGCCAGAGGAGCCCGACCGCCG 2163
 QY 789 GlyLysLeuAlaSerThrProProLysAlaTrpTyrLysAspIleLeuGlnLeuIleGly 808
 DB 2164 GAGGCTGTGCTTCCACCCCAAGGCTGTGTACAGGACATCTGCACCTCATTTGGC 2223
 QY 809 PheAlaAsnLeuProArgValAspGlyTyrCysGlnArgValTyrPyrArgGlyThrThr 828
 DB 2224 TTCCCAACTGCCCCCGGTGGATGACTAGTGAAGCGGTGTGTGTGACGGGCAACAGC 2283
 QY 829 GlucySerGlyCysPheArgSerArgSerArgGlyLysGlnAlaArgGlyLysSerTrp 848
 DB 2284 GAATGCTCAAGCTGTCTCGAGCGGAGCGGAGCGGAGCAGGAGGAGGAGAGAGCTGG 2343
 QY 849 AlaGlyLeuGlnLeuGlyLysLysMetLysSerArgValHisAlaGlnHisAsnArgThr 868
 DB 2344 GCAGGGCTGAGTGAAGCAGAAAGATGAAGACCGGGTGTGATGCCAGCAATCGGAGC 2403
 QY 869 ProArgGlnValAlaThr 875
 DB 2404 CCGCGGAGGTGGAGGCCAGC 2424

RESULT 4
 AAH47791
 ID AAH47791 standard; cDNA; 2349 BP.
 AC AAH47791;
 XX
 XX
 DT 07-JAN-2002 (first entry)
 XX
 DE Novel human protein (NHP) encoding cDNA sequence.


```

QY 494 ProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpProValArgPro 513
DB 1201 CCAGATGAGAGTGTGCTGCACTTGGCCGAGCCACCCCTCATGTTCTGGCCCTGTGGCCCT 1260
QY 514 ArgHisGlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnLeuVal 533
DB 1261 CGACATGAGCGCCCTGCTGCTTGTCAAGAACCCACCTGGCCAGAGCTACACAGATGCTG 1320
QY 534 ValAspArgValGlnAlaGlnAspGlyThrTrpAspValIlePheLeuGlyThrAspSer 553
DB 1321 GTGGACCGGGGTGAGGACAGAGATGGAGCTTACGATGCTATTTCTGGGGAGCTGACTCA 1380
QY 554 GlySerValLeuLysValIleAlaLeuGlnAlaGlySerAlaGlnProGluGluVal 573
DB 1381 GGGTCTGTGCTCAAAAGTCAATGCTTCCAGAGCAGGGGGCTGACCTAGACTGAGGAAGT 1440
QY 574 ValLeuGlnGlnLeuGlnValPheLysValProThrProIleThrIleMetGluIleSer 593
DB 1441 GTTCTGGAGAGAGCTCCAGGTGTTTAAGTGCCAAACACTATCACCAAAATGAGATCTCT 1500
QY 594 ValLysArgGlnMetLeuLysValGlySerArgLeuGlyValAlaGlnLeuArgLeuHis 613
DB 1501 GTCAAAAGCAATGCTATACGTGGGCTCTCGGCTGGGTGTGGCCAGCTGGCGTGCAC 1560
QY 614 GlnCysGluThrTyrGlyThrAlaCysAlaGluCysCysLeuAlaArgAspProTyrCys 633
DB 1561 CAATGTGACACTTACGGCACTGCTGTCAGAGTGTGCTGGCCGGAGACCATACTCT 1620
QY 634 AlaTrpAspGlyAlaSerCysThrHisTyrArgProSerLeuGlyLysArgArgPheArg 653
DB 1621 GCGTGGAGAGTGTGCTCTGCTGACCACTACCCCGCCAGCTTGGCAAGCGCGGTTCCGC 1680
QY 654 ArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCysLeuGlyGlnSerGlnGlnGlu 673
DB 1681 CGGCAAGCACTCCGGCAGCGCAACCTGCTGAGTGGCTGGCGCAGAGCGAGGAAGA 1740
QY 674 GluAlaValGlyLeuValAlaAlaThrMetValTyrGlyThrGlnHisAsnSerThrPhe 693
DB 1741 GAGGAGAGTGGAGCTTGTGGCAGCCACCATGCTCTACGGCAGCAGCAGCAAAVAGCACTTC 1800
QY 694 LeuGlnCysLeuProLysSerProGlnAlaAlaValArgTrpLeuGlnArgProGly 713
DB 1801 CTGGAGTCCCTGCCCACTGCCARCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
QY 714 AspGlnGlyProAspGlnValLysThrAspGlnValLysThrGlnArgGlyLeu 733
DB 1861 GATGAGGGGCTGACAGAGTGAAGAGCAGCAGAGCTTGTGACACAGCAGGAGGGGCTG 1920
QY 734 LeuPheArgArgLeuSerArgPheAspAlaGlyThrTyrThrCysThrThrLeuGlnHis 753
DB 1921 CTGTTCGCGCAGGCTTAGCGCTTGTGATCGGGGCACTACCTGACACCACTCTGGAGCAT 1980
QY 754 GlyPheSerGlnThrValAlaArgLeuAlaLeuValValIleValAlaSerGlnLeuAsp 773
DB 1981 GGGTCTTCCCGAGCTGTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
QY 774 AsnLeuPheProGluProGluProGluProGluProAlaArgGlyGlyLeuAlaSer 793
DB 2041 AACCTGTTCCCTCGGAGCCAAAGCCAGAGAGCCCGCCAGCGGGGCTGCTGCTTCC 2100
QY 794 ThrProProLysAlaTrpTyrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuPro 813
DB 2101 ACCCCACCCAGAGGCTGTGTCAGAGACATCTGACGCTCATTTGCTTGGCCACCTGGCC 2160
QY 814 ArgValAspGlnLysCysGlnArgValTrpCysArgGlyThrTrpGlnCysSerGlyCys 833
DB 2161 CCGGTGATAGTACTGTGAGCGGCTGTGGTGGGAGGAGCAGCAGGATGCTCAGGCTGC 2220
QY 834 PheArgSerArgSerArgGlyLysGlnAlaArgGlyLysSerTrpAlaGlyLeuGlnLeu 853
DB 2221 TTCGCGAGCGGAGCGGCGGAGCAGCAGGAGGCGCAAGAGCTGGGAGGGCTGACTCA 2280

```

```

QY 854 GlyLysLysMetLysSerArgValHisAlaGlnHisAsnArgThrProArgGluValGlu 873
DB 2281 GCGAAGAGATGAAGAGCCGGGTGTCATGCCAGACAAATGAGAGCGCCGGAGAGTGGAG 2340
QY 874 AlaThr 875
DB 2341 GCCACG 2346

RESULT 5
AAFA90250
ID AAF90250 standard; DNA: 2340 BP.
XX
AC AAF90250;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a semaphorin polypeptide designated ZSMF-16.
XX
KW Human; semaphorin; ZSMF-16; neurite growth; neurite outgrowth;
KW T lymphocyte suppressor; cancer; stroke; brain damage; paralysis;
KW spinal injury; neurodegenerative disease; amyotrophic lateral sclerosis;
KW Alzheimer's disease; Huntington's disease; Parkinson's disease;
KW peripheral neuropathy; demyelinating disease; multiple sclerosis;
KW immunosuppression; autoimmune disease; insulin dependent diabetes;
KW rheumatoid arthritis; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT 1..2340
FT /tag= a
FT /product= "semaphorin"
FT sig_peptide
FT 1..66
FT /tag= b
FT /note= "secretory signal sequence"
FT mat_peptide
FT 67..2340
FT /tag= c

WO200140278-A2.
XX
PN 07-JUN-2001.
XX
PD 06-DEC-2000; 2000WO-US3116.
XX
PE 06-DEC-1999; 99US-045560.
XX
PR (ZYMO ) ZYMOGENETICS INC.
XX
PA
XX
PI Holloway JL, Foley KP;
XX
DR WPI: 2001-374784/39.
XX
DR P-PSDB: AAB84219.
XX
PT Novel human semaphorin polypeptide, ZSMF-16, useful for treating
PT peripheral neuropathies Alzheimer's and Huntington's disease and
PT polynucleotide encoding ZSMF-16 useful for detecting genetic
PT abnormality and cancer
XX
PS Claim 2; Page 116-121; 124pp; English.
XX
XX
CC The present sequence encodes a human semaphorin polypeptide, designated
CC ZSMF-16. ZSMF-16 is a neurite growth and development modulator. It also
CC enhances spinal cord and sensory neurite outgrowth and patterning, and
CC is involved in the activation and regulation of T lymphocytes suppressor.
CC ZSMF-16 is useful for detecting a genetic abnormality or cancer. ZSMF-16
CC polynucleotide probes can be used to detect 3p21 loss, trisomy,
CC duplication or translocation associated with mammary tumour tissue,
CC breast tumour, liver, small intestine, bone cancers, etc.. ZSMF-16 can
CC be used to modulate neurite growth and development and demarcate nervous
CC system structures. ZSMF-16 are also useful for regenerating and directing
CC neurite outgrowth following strokes, brain damage caused by head
CC injuries, paralysis caused by spinal injuries, and for treating
CC neurodegenerative diseases such as amyotrophic lateral sclerosis,

```


1798 TGGCTCCCAAGTCTCCAGGCGTCTGCTGCTTTCAGAGGCCAGGAGATGAG 1857
 716 GYProAspGlnValIysTrpAspGluArgValLeuHisThrGluArgGlyLeuLeuPhe 735
 1858 GGGCTTACAGGAGTGAAGAGGAGGAGGAGTCTTGCACGAGGAGGGGCTGCTTCC 1917
 736 ArgArgLeuSerArgPheAspAlaGlyThrTyrThrCysThrThrLeuGluHisGlyPhe 755
 1918 CCAGAGCTTAGCCGTTTCATGCTGGGACCTTACACCTGCACCACTCTGAGCATGGCTTC 1977
 756 SerGlnThrValValArgLeuAlaLeuValValIleValAlaSerGlnLeuAspAsnLeu 775
 1978 TCCCAAGACTGTGGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2037
 776 PheProGlnGluProLysProGluGluProProAlaArgGlyGlyLeuAlaSerTrpPro 795
 2038 TCCCTCCGAGAGCAAGCAGAGGAGGCCGCCAGGCCGGGAGGCTGCTTCCACCCCA 2097
 796 ProLysAlaTrpTyrTyrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuProArgVal 815
 2098 CCCAAGGCTGTGACAAAGACATCTGCAGCTCATGCTGCTGCCCAACCTGCCCGGGGTG 2157
 816 AspGluTyrCysGluArgValTrpCysArgGlyThrThrGluCysSerGlyCysPheArg 835
 2158 GATGAGTACTGTGAGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2217
 836 SerArgSerArgGlyLysGlnAlaArgGlyLysSerTrpAlaGlyLeuGluGlyLys 855
 2218 ACCCGAGAGCGGGGCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2277
 856 LysMetLysSerArgValHisAlaGluHisAsnArgThrProArgGluValGluAlaThr 875
 2278 AAGATGAGAGGCGGGGTGATGCTGCCAGCAGCATCGAGCGCCCGGAGGTGGAGGCGCAG 2337
 RESULT 6
 AAF90251
 ID AAF90251 standard; DNA; 2337 BP.
 AC AAF90251;
 XX
 XX AAF90251;
 XX
 XX 06-AUG-2001 (first entry)
 DT
 XX
 XX Nucleotide sequence of a semaphorin polypeptide designated ZSMF-16.
 DE
 XX
 XX Human; semaphorin; ZSMF-16; neurite growth; neurite outgrowth;
 KW
 KW T lymphocyte suppressor; cancer; stroke; brain damage; paralysis;
 KW spinal injury; neurodegenerative disease; amyotrophic lateral sclerosis;
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease;
 KW peripheral neuropathy; demyelinating disease; multiple sclerosis;
 KW immunosuppression; autoimmune disease; insulin dependent diabetes;
 KW rheumatoid arthritis; ss.
 KW
 XX
 XX Homo sapiens.
 OS
 OS Synthetic.
 XX
 XX MO200140278-A2.
 XX
 XX 07-JUN-2001.
 XX
 XX 06-DEC-2000; 2000WO-US33116.
 XX
 XX 06-DEC-1999; 99US-0455560.
 XX
 XX (ZYMO) ZYMOGENETICS INC.
 XX
 XX Holloway JL, Foley KP;
 PI
 XX WPI, 2001-374784/39.
 DR
 XX
 XX Novel human semaphorin polypeptide, ZSMF-16, useful for treating
 PT peripheral neuropathies Alzheimer's and Huntington's disease and
 PT polynucleotide encoding ZSMF-16 useful for detecting genetic

PT abnormality and cancer
 XX
 PS Claim 3: Page 123-124; 124pp: English.
 XX
 XX The present is a degenerate sequence encoding a semaphorin, designated
 CC ZSMF-16. ZSMF-16 is a neurite growth and development modulator. It also
 CC enhances spinal cord and sensory neurite outgrowth and patterning, and
 CC is involved in the activation and regulation of T lymphocyte suppressor.
 CC ZSMF-16 is useful for detecting a genetic abnormality or cancer. ZSMF-16
 CC polynucleotide probes can be used to detect 3p21 loss, trisomy,
 CC duplication or translocation associated with mammary tumour tissue,
 CC breast tumour, liver, small intestine, bone cancers, etc.. ZSMF-16 can
 CC be used to modulate neurite growth and development and demarcate nervous
 CC system structures. ZSMF-16 are also useful for regenerating and directing
 CC neurite outgrowth following strokes, brain damage caused by head
 CC injuries, paralysis caused by spinal injuries, and for treating
 CC neurodegenerative diseases such as amyotrophic lateral sclerosis,
 CC Alzheimer's disease, Huntington's disease, Parkinson's disease and
 CC peripheral neuropathies, or demyelinating diseases e.g., multiple
 CC sclerosis. ZSMF-16 also acts as a mediator of immunosuppression,
 CC and thus useful for diagnosing and treating autoimmune diseases such as
 CC insulin dependent diabetes, rheumatoid arthritis, and multiple
 CC sclerosis. It can also be used as an anti-inflammatory for inhibition
 CC of antigen in humoral and cellular immunity and for immunosuppression
 CC in graft and organ transplants.
 CC
 XX
 SQ Sequence 2337 BP; 308 A; 273 C; 468 G; 293 T; 995 other;
 Alignment Scores:
 Pred. No.: 1, 1e-161 Length: 2337
 Score: 3102.50 Matches: 573
 Percent Similarity: 71.75% Conservative: 1
 Best Local Similarity: 71.62% Mismatches: 187
 Query Match: 65.37% Indels: 39
 DB: 22 Gaps: 2
 US-09-813-290-2 (1-875) x AAF90251 (1-2337)
 QY 94 MetaLaprosSerAlaTrpAlaIleCysTrpLeuGlyGlyLeuLeuLeuHisGlyGly 113
 DB 1 ATGGCNCNCMMNSNCNTGGCCNATHTGTGTYTNGNGNNTNTNTNCAYGNGN 60
 QY 114 SerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTyrArgAspLeu 133
 DB 61 WSNWSNGNCNCMMNSNCNCNGNCNCNMGNTNMGNNTNMSNTNAYMGNGNGCN 120
 QY 134 Leu-----Ser 135
 DB 121 ATGGTNGNARCCMMNSNMCNATGTGATGARACNTTWSNMNGNTAYTNTNMSN 180
 QY 136 AlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetLys 155
 DB 181 GCNAAYMGNSNMCNATHTTNTTNGNCNCNCARGGWMSNYTNAAYTNCARCCNATGTAY 240
 QY 156 LeuAspGluTyrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArg 175
 DB 241 YTNAGYAGTATMGAGAYMGANTNTTNTTNGNGNNTNGAAGCAYTNTAYMSNTNMCN 300
 QY 176 LeuAspGlnAlaTrpProAspProArgGluValLeuTyrProGlnProGlyGlnArg 195
 DB 301 YTNAGYAGCAGCTGGCCNGAYCCNMNGAR----- 330
 QY 196 GluGluCysValArgLysGlyArgAspProLeuThrGluCysAlaAsnPheValArgVal 215
 DB 331 -----ACNGARTGTGCNNAAYTNTGNTNGNGTN 357
 QY 216 LeuGlnProHisAsnArgThrHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThr 235
 DB 358 YTNCAACCCNCAAYAAAGNACAYTNTNTGNCNTGYGAGNAGNGNCNTTTCARCCNMCN 417
 QY 236 CysAlaLeuIleThrValGlyHisArgGlyHisValLeuHisLeuGlnProGlySer 255
 DB 418 TGYGCTNTNATNACNGTNGNCATYMGNGNGARCAAYTNTNCAYTNGARCCNCGNMSN 477

```

OY 256 ValGluSerGlyArgGlyArgCysProHisGluProSerArgProPheAlaSerThrPhe 275
Db 478 GTNGARWMSGNGMNGMNGMTGTCACCAICARCCNMSNMNGCCNTTTCGCGWSMACNTTY 537
OY 276 IleAspGlyGluLeuTyThrGlyLeuThrAlaAspPheLeuGlyArgGluAlaMetIle 295
Db 538 ATHGAYGNGARNTTAVACNGAYTNACNGAYTTCGNGMNGMNGGARGCATGATH 597
OY 296 PheArgSerGlyGlyProArgProAlaLeuArgSerAspSerAspGlnSerLeuHis 315
Db 538 TTYMGMSWMSGNGMNGCCMNGCCNTTNGMNSNGAYWSNGAYCARMSTYNTWTCAY 657
OY 316 AspProArgPheValMetAlaAlaArgIleProGluAsnSerAspGlnAspAsnAspLys 335
Db 658 GAYCCNMGTTCGTATGACGCGCCNCGNATGCCNGARAAWMSGAYCARGAAYAAAYCAAR 717
OY 336 ValTyrrPhePhePheSerGluThrValProSerProAspGlyGlySerAsnHisValThr 355
Db 718 GTNTATYTTTTCGTTWMSGARACNCCNCCMNSCNGAYGGGWSMAAICATGTCNACN 777
OY 356 ValSerArgValGlyArgValCysValAsnAspAlaGlyGlyArgValLeuValAsn 375
Db 778 GTNWSMNGMGTNGMNGMNTNGTGTNAAYGCGNGCGNGCARMGNTNTNGTNAAY 837
OY 376 LysTrpSerThrPheLeuLysAlaArgLeuValCysSerValProGlyProGlyGlyAla 395
Db 838 AATGCGMSNACNTTTCATNARCGNMNGNTNGTNGTWSNGTNGCNGCGNGCGNGCGNGCN 897
OY 396 GluThrHisPheAspGluLeuGluAspValPheLeuLeuTrpProLysAlaGlyLysSer 415
Db 898 GARACCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 957
OY 416 LeuGluValTyrrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCys 435
Db 958 YTGAGAGTNTATTCNTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1017
OY 436 ValTyrrHisMetAlaAspIleTrpGluValPheAsnGlyProPheAlaHisArgAspGly 455
Db 1018 GTNTATYATTCGCGNGAYATTCGAGTNTTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1077
OY 456 ProGlnHisGlnTrpGlyProTyrrGlyGlyLysValProPheProArgProGlyValCys 475
Db 1078 CCNCAKATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1137
OY 476 ProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTyrrProAsp 495
Db 1138 CCNMSNAARATGACNGCNCARCCNGMNGMNGCCNTTTCGNGMSNACNARATATCCNGAY 1197
OY 496 GluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpProValArgProAlaHis 515
Db 1198 GARGTNTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1257
OY 516 GlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleValValAsp 535
Db 1258 GGNMGCCNGTNTNGTNAARACNCAAYTTCNCAARATTCATTCATTCATTCATTCATTCATTCAT 1317
OY 536 ArgValGluAlaGluAspGlyThrTyrrAspValIlePheLeuGlyThrAspSerGlySer 545
Db 1318 MONGTNGARGCGARGAGVAGNACNTATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
OY 556 ValLeuLysValIleAlaLeuGlnAlaGlyLysSerAlaGluProGluGluValValLeu 575
Db 1378 GTNTATYATTCGTTWMSGARACNCCNCCMNSCNGAYGGGWSMAAICATGTCNACN 1437
OY 576 GluGluLeuGluValPheLysValProThrProIleThrGluMetGluIleSerValLys 595
Db 1438 GARGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1497
OY 596 ArgGluMetLeuTyrrValGlySerArgLeuGlyValAlaGlnLeuArgLeuHisGlnCys 615
Db 1498 MGCAATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1557

```

```

OY 616 GluThrTyrrGlyThrAlaCysAlaGluCysCysLeuAlaArgAspProTyrrCysAlaTrp 635
Db 1558 GARACNTATGAGNACNCCNTTCGCGNARTrpGlyTTCGCGMNGMNGAACCNTATTCGCGNCG 1617
OY 636 AspGlyAlaSerCysThrHisTyrrArgProSerLeuGlyLysArgArgPheArgGln 655
Db 1618 GAYGNGCNSMNTTCYACNCAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1677
OY 656 AspIleArgHisGlnLysProAlaLeuGlnCysLeuGlyGlnSerGlnGluGluAla 675
Db 1678 GAYATTHGNCAYGGAAYTCGCGNTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1737
OY 676 ValGlyLeuValAlaAlaThrMetValTyrrGlyThrGluHisAsnSerThrPheLeuGlu 695
Db 1738 GTNGANTTNGTNGCNGACNATGTTATGAGNACNARCAAYAAWMSNACNTTTCGTTGAR 1797
OY 696 CysLeuProLysSerProGlnAlaAlaValArgTrpLeuLeuGlnArgProGlyAspGlu 715
Db 1798 TGYTTCNCAARMSNCCNCAKCGCGTNGMNGTTCGTTTCATTCATTCATTCATTCATTCATTCAT 1857
OY 716 GlyProAspGluValLysThrAspGluArgValLeuHisThrGluArgGlyLeuPhe 735
Db 1858 GCGNCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1917
OY 736 ArgArgLeuSerArgPheAspAlaGlyThrTyrrCysThrThrLeuGluHisGlyPhe 755
Db 1918 MGNMGNTTNGMNGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1977
OY 756 SerGlnTrpValAlaArgLeuAlaLeuValIleValAlaSerGlnLeuAspAsnLeu 775
Db 1978 WSNCAKACNCTGTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2037
OY 776 PheProProGluProLysProGluGluProProAlaArgGlyGlyLeuAlaSerThrPro 795
Db 2038 TTCNCCNGACNCAARACNCAKCGNCCNCCMNGMNGGNGGTYTTCGWSNACNCCN 2097
OY 796 ProLysAlaTrpTyrrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuProArgVal 815
Db 2098 CCNARCGTGTATTAARATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2157
OY 816 AspGluTyrrCysGluArgValTrpCysArgGlyThrThrGluCysSerGlyCysPheArg 835
Db 2158 GAYGATATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2217
OY 836 SerArgSerArgGlyLysGlnAlaArgGlyLysSerTrpAlaGlyLeuGluLeuGlyLys 855
Db 2218 WSNMGMSWMSGNGMNGCNCARCCNCGMNGMNGGNAARWSNTGGCGNGTYTTCATTCATTCATTCATTCAT 2277
OY 856 LysMetLysSerArgValHisAlaGlnHisAsnArgThrProArgGluValGluAlaThr 875
Db 2278 AARATGAARMSNMGNTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 2337

RESULT 7
AAV35367 standard; cDNA: 2898 BP.
AC AAV35367;
XX 01-OCT-1998 (first entry)
XX Human semaphorin encoding cDNA.
XX DE
XX Human; semaphorin; diagnosis; nerve disease; nerve growth inhibitor;
XX KW neurological disease; atopic skin inflammation; autoimmune disease;
XX KW pain; ds.
XX Homo sapiens.
XX OS
XX FH key Location/Qualifiers
XX CDS 370..2697
XX FT /*tag= a
XX FT /product= "semaphorin"
XX

```



```

OY 550 YTHASPSERGLYSERVALLEULYSVALLEALALEUGLNLAAGLYGYSERVALLEULPR 570
DB 1731 GACGACACACAGCAATTCGCTGCAAGATATCAACATTTACAAACCAAGAAAGAGTGGAT 1790
OY 570 OGLUGLVALLEUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 590
DB 1791 GGAGGAGTCAATTCATAGAGCACTCAATTCATAGAGTCAATTCATAGAGTCAATTCAT 1850
OY 550 TGLULESERVALLYSARGLINMETLEUTYRVALGYSERARGLEUGLYVALAGLIND 610
DB 1851 GGAATTTCTTCACAAAGACACACACACCTTTACATTTGATCAGGCTGCTGCTGCAACAGT 1910
OY 610 UARGLEUHLISGLNYSGLURTHYRGLYTHRALCYSAALAGLYGYSLEUALAARGAS 630
DB 1911 CAGATTCATCATCCTCGACATGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1970
OY 630 PPROTYRCYSALATTPASBGLYALASERCYSRTHIRHSTYRARGPRO-----SERLEUGL 648
DB 1971 CCCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2030
OY 648 YLYSARARGPHEARGARGGLNAPLLEARGHISGLYASPROALALEUGLINCYLEUGL 668
DB 2031 AAGAGAGAGTTCGCCAGCAGCAGCAGCTTGGCATGGCAACGCCCAACAGTGTGG 2090
OY 668 YGLNSERGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 688
DB 2091 ACAGCAATTTGCTGAGACGCCGCTGGACAGACTGAAGAGGCTGCTGCTGCTGCTGCTGCT 2150
OY 688 UHLSASNSERPHLEUGLUGLYLEUPROLYSLEUPROLYSLEUPROLYSLEUPROLYS 708
DB 2151 GACCAACAGAGTCTCTTGGATATGACCCGCCATCATCTCAACAGCAAAAGTCACTGCTGT 2210
OY 708 ULEUGLARGPROGLYASPLUGLYPROASPLUGLYVALYSTHASPGLUARGYALLEUHL 728
DB 2211 TGTACAGAAAGGAGCAGCTGACGAGTCAAGAAAGAGTCAAGAGTCAAGAGTCAAG 2270
OY 728 STRHGLUARGLYLEULEUHEARGHARGLEUSERARGPHEASPLAAGLYTHRYRTHRYCY 748
DB 2271 GATGAGACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2330
OY 748 STRHTRLEUGLUNHISGLYHESERGLNTHVALVALARGLEUALALEUVALLEUHL 768
DB 2331 CCAGACAGTGAACACACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2390
OY 768 LALASERGLINLEUASPLINLEUPROPROGLUINPROLYSLEUGLUGLUGLUGLUGL 784
DB 2391 AGAGCATTAAGGTGAGGCGCATGTTTCATTAAGGACCATGAAAGAGAAACATCAACAT 2450
OY 785 -----PROPROALARGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 802
DB 2451 GCCCTGCCCTCCCTTAAGCGGTATGCTCAAGGAGCAAAACCG-----TGTACAAAGA 2504
OY 802 PILEUGLUNLEUILLLEGLYPHEALASINLEUPROARGYVALASPLUGLYRYSGLUARGVA 822
DB 2505 ATTCTTGACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2564
OY 822 LTRPCYSARGLYHTRHGLUCYSERGLYCYSPHEARGSERARGSERARGLYLYSG 842
DB 2565 GTGGGTCACAGT-----AAGAAAGAGAAAGAAAGCTTAAGT 2600
OY 842 NALARGLYLYSSETRPALAGLYLEUGLUGLUGLYLYSGLYSMEULYSSEARGYALHI 862
DB 2601 GTCCTCCCTCAAGTGAAGTATGCAACCCCGAGAAAGAGGCTTGTCTTAA----- 2655
OY 862 SALAGLUNHISASARGTHPROARG 870
DB 2656 -GCTGACACTTCGCCGCTGCCAGG 2679

```

RESULT 8
AA228469 standard; DNA; 4460 BP.
XX
AC AA228469;

```

XX 05-JAN-2000 (first entry)
DT Mouse semaphorin H (Sema H) polynucleotide sequence.
DE
XX
XX Semaphorin H; Sema H; Sema Hv; collapsin; bone structure formation;
KW metastasis; cancer; antibody; drug screen; ss.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
FH 609..2937
FT CDS /*tag= a
FT /product= M.Sema_H
FT /transl_except=(Pos:2452..2454, aa:Tyr)
FT
XX
XX WO947671-A2.
XX
XX 23-SEP-1999.
XX
XX 12-MAR-1999; 99WO-1B00495.
XX
XX 13-MAR-1998; 98US-0077997.
XX
XX (LURA/) LUKANIDIN E M.
XX (CHRT/) CHRISTENSEN C R L.
XX
XX Lukanidin EM, Christensen CRL;
XX
XX WPI; 1999-590975/50.
XX P-P-SDS; AA228469.
XX
XX New polypeptides and polynucleotides, useful in diagnosis and treatment
XX of metastatic cancer
XX
XX Claim 1; Page 78-85; 95pp; English.
XX
XX This is the mouse Semaphorin-H polynucleotide sequence (Sema H). The
XX Semaphorin/collapsin family of molecules are characterised by a unique
XX and highly conserved motif, within a 500 amino acid semaphorin domain.
XX Some semaphorins exhibit inhibitory or repulsive functions in a neuronal
XX context, and functions in bone structure formation are also implicated.
XX It is thought that Sema H plays a role in metastasis. The invention uses
XX the mouse Sema H gene in the diagnosis of metastatic cancer. Semaphorin
XX polynucleotides and polypeptides are used in the methods of the
XX invention, the polypeptides are useful for determining the metastatic
XX potential of cells, by detecting their expression in biological samples.
XX Antibodies specific for Sema H, are also useful therapeutically in
XX inhibiting Sema-H polypeptide activity and therefore metastasis, and for
XX purifying the polypeptides. Metastasis may also be inhibited by
XX inhibiting the biological activity of the polypeptide using e.g. a small
XX molecule inhibitor or a Semaphorin-H ligand (or fragment). The
XX polynucleotides can also be used to inhibit polypeptide expression in
XX cells using known antisense technology e.g. to prevent metastasis of
XX cancer cells. They can be used to detect and quantify Sema-H mRNA levels
XX in cells. The polypeptides, fusion proteins, multimeric proteins,
XX antibodies or antisense oligonucleotides can be included in
XX pharmaceutical compositions. The polynucleotides can be used to isolate
XX similar sequences from other species and to produce mammalian cell lines
XX and tumours with known metastatic potential, useful in anti-metastatic
XX drug screening.
XX
XX Sequence 4460 BP; 1288 A; 950 C; 1023 G; 1199 T; 0 other;
SQ

```

Alignment Scores:

Pred. No.: 6,27e-104 Length: 4460
Score: 2056.00 Matches: 431
Percent Similarity: 60.93% Conservative: 135
Best Local Similarity: 46.39% Mismatches: 268
Query Match: 43.32% Indels: 95
DB: 20 Gaps: 18

US-09-813-290-2 (1-875) x AA228469 (1-4460)

[illegible]


```
QY 80 -----GlySerAlaGluArgGlnArgCysProGlnPheProSer 93
Db 565 TCCTAAAGTGTGGTGCAGGGGGCCAGAGAGAGCCAGCACC-----ACC 609
QY 94 MetAlaProSerAlaIlePheCysTrpLeuLeuGlyLeuLeuHisGly 113
Db 610 ATGGACCCGGCCGGACACATCTCAGCTTGGCTGGGTGACACTGTGGAACCTGG 669
QY 114 SerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTrpArgAspLeu 133
Db 670 ACCCCAGGTCTCTGGGGAGACCCCTCTACCCGAGGCTACCGCTGTGCATTAAGAAGCTT 729
QY 134 LeuSerAlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnGlnAla 153
Db 730 TTCGAACCTGATGACTTTCATATTTCAAGCCCTTGGATTTCCTGATCTCATACA 789
QY 154 MetTrpLeuAspGluTrpArgAspArgLeuPheLeuGlyLeuLeuAspAlaIlePheGln 173
Db 790 ATGCTGCTGGATGGATATCAAGAAGCGCTCTTGTGGAGGACAGACCTTGTCTATTCC 849
QY 174 LeuArgLeuAspGlnAlaIlePheProAspProArgGluValLeuTrpProGlnProGly 193
Db 850 CTGAACCTTGGACAGAGTGAAGGAGCGCTACAGAGATATATCTGCCGAGACAGCACTA 909
QY 194 GlnArgGlnGluCysValArgGlyArgAspProLeuThrGlnCysAlaAsnPheVal 213
Db 910 AACGTAGAAAGATGCAATTAAGAAAGAAAGAC---GCAATGATGCTGCCAATATTATATC 966
QY 214 ArgValLeuGlnProHisAsnArgTrpHisLeuLeuAlaCysGlyTrpGlyAlaPheGln 233
Db 967 CGGGTTTTCATCATCAACAGACACACCTTCTGACCTGTGCTACTGCTGCTTGTGAT 1026
QY 234 ProThrCysAlaLeuIleThrValGlyHisArgGlyGlnHisValLeu---HisLeuGln 252
Db 1027 CCACACTGTGCTTCATCAAGATCGGGCCACCATTCAGAGAACCCCTGTTTACCTGGAG 1086
QY 253 ProGlySerValGlnSerGlyArgGlyArgCysProHisLeuProSerArgProPheAla 272
Db 1087 TCACACAGATCTAGAGAGAGAGAGGAGATGTCCTTTGACCCCACTCTCTCTTGTG 1146
QY 273 SerThrPheIleAspGlyLeuLeuTrpGlyLeuThrAlaAspPheLeuGlyArgGln 292
Db 1147 TCCACCTAGTTGGATAGCTGTTTGTGCGACTGTACAGCTATATTTGGCCAGAGAC 1206
QY 293 AlaMetIlePheArgSerGlyGlyProArgProAlaLeuArgSerAsp---SerAspGln 311
Db 1207 TCGGCAATCTCGGACGATGGGGAAGTTAGGCCATATTGCGACTGACATGACATGAG 1266
QY 312 SerLeuLeuHisAspProArgPheValMetAlaAlaArgIleProGluAsnSerAspGln 331
Db 1267 CGGCTCTCGAAGAACCAAAATTGTAGGTTCATATATGATTTCTCATACGAAGACGA 1326
QY 332 AspAsnAspLysValTrpPhePhePheSerGluThrValProSerProAspGlySer 351
Db 1327 GATGACAAACAAATGATCTTTTCTTACTGAGAAAGCCCTGGAGCGGAGAACACAGCC 1386
QY 352 AsnHisValThrValSerArgValGlyArgValCysValAsnAspAlaGlyGlnArg 371
Db 1387 ---CACACAGATCTTACACCCGAGGTGGGCTGTGCGTAATGACATGGAGAGACAGAA 1443
QY 372 ValLeuValAsnLysTrpSerThrPheLeuLysAlaArgLeuValCysSerValProGly 391
Db 1444 ATCCGTGTCGAACAGATGAGACATTTCTTAAGCGCGGCTGTGTGCTCATGCGCGGGA 1503
QY 392 ProGlyGlyAlaGlnThrHisPheAspGlnLeuGluAspValPheLeuLeuTrpProLys 411
Db 1504 ATGAATGGAAATCGACATACTTGTGACGAACCTAGAGGATGTTTACTGCCGACCA 1563
QY 412 AlaGlyLysSerLeuGluValTrpAlaLeuPheSerThrValSerAlaValPheGlnGly 431
Db 1564 GATCTTAAGATCCAGTATATTTGACTGTTTAATACACCAATATATATTAGAGGC 1623
QY 432 PheAlaValCysValTrpHisMetAlaAspIleTrpGluValPheAsnGlyProPheAla 451
Db 1624 CATGCTGATGTGTATACATATGCAAGTATCCGGGAAGCCTTAATAGGCCATATGCT 1683
QY 452 HisArgAspGlyProGlnHisGlnTrpGlyProTrpGlyGlyValProPheProArg 471
Db 1684 CATAAGAAGGCCCTGATATACACCTGCTACTATATGAAGAAAGTCCCTACCCAAAG 1743
QY 472 ProGlyValCysProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLys 491
Db 1744 CCGTGTCTCTGCGCAGCAAGTAAC-----GGAGCAGATATGAGAACCAACAA 1794
QY 492 AspTrpProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpProVal 511
Db 1795 AGATTACCCGATGACGCCATCCGCTGCGAAGATGCACTCTCTATATGTATACCCCAT 1854
QY 512 ArgProArgHisGlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGln 531
Db 1855 AACCTGTCATTAATAAAACCAATCTGGTAAACAGATGAAATACCACTGACGCA 1914
QY 532 IleValAlaAspArgValGluAlaGluAspGlyThrTrpAspValIlePheLeuGlyThr 551
Db 1915 CTGCCCGTGAATCGGGTGAAGCGAGAGATGCGCATGATGACTTATATTATGGCACA 1974
QY 552 AspSerGlySerVal---LeuLysValIleAlaLeuGlnAlaGlySerAlaGluPro 570
Db 1975 GACACAGGAATTTGCTGCTGAAGTATATCAATTTACACCAAGAAAGAGCTGATG 2034
QY 571 GluGluValValLeuGlnGluLeuGlnValPheLysValProThrProIleThrGluMet 590
Db 2035 GAGGAAGTCTATCTAGAGAACTTCAATATTTCAAGATGCCAGCCCTATCATTTTATG 2094
QY 591 GluIleSerValLysArgGlnMetLeuTrpValGlySerArgLeuGlyValAlaGlnLeu 610
Db 2095 GAATTTCTTCAAGAAACACACACCTTACATGTGATGACCTCTGCTGGCACAAGTC 2154
QY 611 ArgLeuHisGlnCysGluThrTrpGlyThrAlaCysAlaGlnCysCysLeuAlaArgAsp 630
Db 2155 AGATTCATCATCTCGACATGATGAGCATGTCCTGTGCTACTGTCTGCTGCTGCGAGAC 2214
QY 631 ProThrCysAlaIleTrpAspGlyAlaSerCysThrHisTrpArgProSerLeuGly----- 648
Db 2215 CCGTACTCTCTCGGATGGATCATATCTGCTCCAGGTATACCAACAGAGCTGCACAGAA 2274
QY 649 LysArgArgPheArgArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCysLeuGly 668
Db 2275 AAGAGGAGTTCGCGACGACGACGACGTCCTGCGCATGGCAGACCCGCCAACAGTCTTGA 2334
QY 669 GlnSerGlnGluGlnAlaValGlyLeuValAlaAlaThrMetValTrpGlyThrGlu 688
Db 2335 CAGCAATTTGTTGGAGACGCTTGGACAGACTGAAAGAGAGCTGCTATGCAATAGAG 2394
QY 689 HisAsnSerThrPheLeuGlnCysLeuProLysSerProGlnIleAlaValArgTrpLeu 708
Db 2395 AGCAACAGTACTCTGTGGATGACACCCGCTCTACCTACAAACCAAAAGCATCTGTGTT 2454
QY 709 LeuGlnArgProGlyLysAspGlnGlyProAspGlnValLysThrAspLysArgValLeuHis 728
Db 2455 CTACAGAAAGAGCGACGCTAAGAAAGAAAGAGAGTGAAGCGAGTACAGAGTGTGCAAG 2514
QY 729 ThrGluArgGlyLeuLeuPheArgArgLeuSerArgPheAspAlaGlyThrTrpTrpCys 748
Db 2515 ATGGACTTGGGCTGTCTCTCAAGATGACGCAAGTCAAGATGCAAGGACCTATTTTCTC 2574
QY 749 ThrThrLeuGlnHisGlyLysSerGlnThrValAlaArgLeuAlaLeuValIleVal 768
Db 2575 CAGACAGTGAACACATTTTGTCCATCTACTGTGCGGTAATAATCACTTGGAGGTGTCGA 2634
QY 769 AlaSerGlnLeuAspAsnLeuPheProProGluProLysProGlnGlu----- 784
Db 2635 GAGCATTAAGTGGAGGCGCTTTCATTAAGACCATGAAAGAGAAAGACATCAAAAGATG 2694
QY 785 -----ProProAlaArgGlyGlyLeuAlaSerThrProProLysAlaTrpTrpLysAsp 802
```

Db 2695 CCCTGCCCCCTTAAGCGGTATGTCAGGCGACAAACCG-----TGCTACAGGAA 2748
 QY 803 ILeuGlnLeuIleGlyPheAlaAsn---LeuProArgValaIspGlyTyrCysGluArg 821
 Db 2749 TTCCTTGACGCTGATGGCTACAGCAGCAAGTCCAGAGGTGGAAGATGACTGCCAAAG 2808
 QY 822 ValTTPCysArgGlyThrThiGluCysSerGlyCysPheArgSerArgSerArgIlys 841
 Db 2809 GTGTGGGTGTACAGAT-----AAGACAGGAGAAAAAGCTTAAA 2844
 QY 842 GlnAlaArgGlyLysSerTrrPalaglyLeuGlyLysLysMetLysSerArgVal 861
 Db 2845 ATGCTCTCCCTCAAGTGAAGTATGCCAACCCCGAAGAAAGAGCTTCGCTTAAA--- 2901
 QY 862 HisAlaGlnHisAsnArgThrProArg 870
 Db 2902 ---GCTGAGCACTTCGCGCTGCCAGG 2925
 RESULT 10
 AAH47049
 ID AAH47049 standard; DNA; 2709 BP.
 AAH47049;
 29-OCT-2001 (first entry)
 Semaphorin D cDNA sequence.
 Drug resistance gene; semaphorin D; B94; mel-14 antigen
 24p3; proliferation; maspin; cancer; cytostatic; gene therapy; ss.
 Homo sapiens.
 WO200155455-A2.
 02-AUG-2001.
 31-JAN-2001; 2001WO-US03161.
 31-JAN-2000; 2000US-0179191.
 (MILL-) MILLENNIUM PHARM INC.
 (JINS/) JIN S.
 Jin S;
 WPI; 2001-488799/53.
 Determining if a compound modulates the drug resistance of a cell,
 comprises determining the expression or activity level of a resistance
 sequence in a cell in the presence of the test compound -
 Example 1; Fig 1A-B; 79pp; English.
 The invention relates to a method of determining whether a test compound
 modulates the drug resistance of a cell that comprises determining the
 expression or activity level of resistance genes (e.g. semaphorin D, B94,
 mel-14 antigen, 24p3, proliferation or maspin) in a cell in the presence of
 the test compound, and comparing its expression or activity level in a
 cell without the test compound. The drug resistant sequences are useful
 in identifying drug resistant cells, in screening methods directed to the
 identification of compounds that can modulate the drug resistance of a
 cell type or multiple cell types. An isolated resistance protein can be
 used as an immunogen to generate antibodies that bind the resistance
 protein. Resistance nucleic acids may be inserted into vectors and used
 as gene therapy vectors. An anti-resistance protein antibody may be used
 to isolate a resistance protein, or facilitate the purification of
 natural resistance protein from cells and of recombinantly produced
 resistance protein expressed in host cells. The methods are useful for
 treating a subject having a disorder, such as a drug-resistance cancer,
 characterized by aberrant resistance sequence expression or activity by
 administering to the subject a resistance modulator. The present sequence
 represents a semaphorin cDNA sequence, whose expression was increased in

CC drug resistant EMT6 tumours.
 XX Sequence 2709 BP; 831 A; 560 C; 620 G; 698 T; 0 other;
 SO Alignment Scores:
 Pred. No.: 5.7e-92 Length: 2709
 Score: 1834.00 Matches: 378
 Percent Similarity: 62.19% Conservative: 150
 Best Local Similarity: 44.52% Mismatches: 262
 Query Match: 38.64% Indels: 60
 DB: 22 Gaps: 16
 US-09-813-290-2 (1-875) x AAH47049 (1-2709)
 QY 43 LeuSerSerAlaProLeuProAlaGlnAspTrpValGluProLeuProTyrLysTrp 62
 Db 87 CTCGGCTGTTCCTCCATTCATGTCAGCCAGTCT---ATTCCAGATGTGTTGAAGTTCCTG 143
 QY 63 ProGlyLysArgAlaAsnTyrAsnArgArgProAlaGlyProGlyLysSerAla 82
 Db 144 CCG-----GACATACA 155
 QY 83 GlyArgArgGlnArgCysProGlnPheProSerMetAlaProSerAlaTrpAla----- 100
 Db 156 GGAGAGAGACTAAAGCACCAAGGAGCTACA---GCGTTCGACATGGAGCTGTTAA 212
 QY 101 -----IleCysTrpLeuLeuGlyLysLeuLeuHisGlyLysSerArgProSer 118
 Db 213 CTAGGATTTGCTCTCT---CTTTTCTGGGAGATATTAATACAGCAAGCAATCTCAGAAAT 271
 QY 119 ProGlyProSerValProArgLeuArgLeuSerTyrArgAspLeuLeuSerAlaAsnArg 138
 Db 272 GGAGAGAGCAATGTGCCAAGGCTGAATATCTTACAAAGAAATGTGGAATCAACAAT 331
 QY 139 SerAlaIlePheLeuLeuProGlnGlySerLeuAsnLeuGlnAlaMetTyrLeuAspGlu 158
 Db 332 GTGATCATTTCAATGAGCTTGGCCCAACAGCTCCAGTTATCATACCTTCCTTTGGATGAG 391
 QY 159 TyrArgAspArgLeuPheLeuGlyLysLeuAspAlaLeuTyrSerLeuArgLeuAspGln 178
 Db 392 GAGCGAGATGAGCTGTATGTGGAGCAAGAGATCACATATTTATTCATTCGAGCTGTTAAT 451
 QY 179 AlaTrpProAspProArgGluValLeuTrpProProGlnProGlyGlnArgGluGluCys 198
 Db 452 ATC---AAGATTTTCAAAAGATTTGCTGGCCAGATATCTTACACCAAGAGATGAATGC 508
 QY 199 ValArgLysGlyArgAspProLeuThrGluCysAlaAsnPheValArgValLeuGlnPro 218
 Db 509 AAGTGGCTGGAAGAGCATCTGAAAGATGTGCTAATTTTCATCAGAGTACTTAAGCA 568
 QY 219 HisAsnArgThrHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeu 238
 Db 569 TATATACAGACTCTTGTACGCTGTGGAACGGGCTTTTCATCAATTTGACCTTAC 628
 QY 239 IleThrValGlnHisArgGlyGlu---HisValLeuHisLeuGlnProGlySerValGlu 257
 Db 629 ATTCAATTTGACATCTACCTGAGGACAAATTTTAAAGCTGGGAATCAGATTTTGA 688
 QY 258 SerGlyArgGlyArgCysProHisGluProSerArgProPheAlaSerThrPheLeuAsp 277
 Db 689 AAGCGCGCTGGGAAGAGTCCATGACCCCTTAAGCTGAGAGAGATCCCTTTAATGAT 748
 QY 278 GlyLeuLeuTyrThrGlyLeuThrAlaAspPheLeuGlyArgGluAlaMetIlePheArg 297
 Db 749 GGAGAAATTAATCTGTGAGTCAAGCTGATTTTATGGGCGAGACTTGTGATCTTCCGA 808
 QY 298 SerGlyArgProArgProAlaLeuArgSerAspSer---AspGlnSerLeuLeuHisAsp 316
 Db 809 ACTCTTGGGACCAACCCCAATGAGGACAGAGACATGATTTCCAGGTGGCTCATGAT 868
 QY 317 ProArgPheValMetAlaAlaArgIleProGluAsnSerArgGlnAspAsnArgVal 336
 Db 869 CCAAGATTCATTAGTGGCCACCTGATCTCAGAGAGTGCATTCCTGAGATGACAAAGTA 928


```

QY 589 GlnMetGluLeuSerValIysArgGlnMetLeuValIglySerArgGluValAla 608
DB 1486 GCATGGAGCTTCCACTACACAGCAACAACTATATATGTTGACAGCGCTGGCTTCC 1545
QY 609 GlnLeuArgLeuHisGlnCysGluThrThrGlyThrAlaCysAlaGluCysCysLeuAla 628
DB 1546 CAGCTCCCTTCCACCGCTGATATTTACCGGAAGCGTGTGCTGATGCTTCCCTCC 1605
QY 629 ArgAspProGlyCysAlaIleProGlyAlaSerCysThrHisThrArgProSerLeuGly 648
DB 1606 CGAGACCTTACTGCTGTTGGATGTTCTGCAATGTTCTGCTATTTTCCACG---GCA 1662
QY 649 LysArgArgPheArgArgGlnAspIleArgHisGlnAspProAlaLeuGlnCysLeu--- 667
DB 1663 AAGAGACGACAGACAGACAGATATAAGAAATGAGACCCACGACCTACATGTTCCAGAC 1722
QY 668 -----GlyGlnSerGlnGluGluAlaValGlyLeuValAla 680
DB 1723 TTACACCATATATATCCATGCGCACAGCCCTGAGAGAGA----- 1764
QY 681 AlaThrMetValTyrGlyThrGlnHisAsnSerThrPheLeuGlnCysLeuProLysSer 700
DB 1765 -----ATCATCTATGCTGTAGAGAAATAGTACACATTTTGGATGCACTCCAGAGTCG 1818
QY 701 ProGlnAlaAlaValAlaArgTyrPheLeuGlnArgProGlyAspGluGlyProAspGlnVal 720
DB 1819 CAGAGAGCGCTGCTGTATGTCGAATTCAGAGAGCGGAATGAGAGCAAAAGAGAGATC 1878
QY 721 LysThrAspGluArgValLeuHisThrGlnArgGlyLeuLeuPheArgGluSerArg 740
DB 1879 AGAGTGCATATATATATATATATATATATATATATATATATATATATATATATATAT 1938
QY 741 PheAspAlaGlyThrTyrThrCysThrThrLeuGlnHisGlyPheSerGlnThrValAla 760
DB 1939 AAGGATTCAGCAATATATATATATATATATATATATATATATATATATATATATATAT 1998
QY 761 ArgLeuAlaLeuValValIleValAlaSerGlnLeuAspAsnLeuPheProGluPro 780
DB 1999 AAGGTAACTCGAAGATCATGTCACACAGACATTTGGAAAGAACTCTTCAATAAGATGAT 2058
QY 781 LysProGluGluProProAlaArgGlyGlyLeuAlaSerThrProPro-----LysAla 798
DB 2059 GATGAGATGCTCTAAGACCAAGAAATGTCATATGATGACACTTACCCAGAAAGTC 2118
QY 799 TrpTyrLysAspIleLeuGlnLeuIleGlyPheAlaAsnLeuProArgValAspGluTyr 818
DB 2119 TGGTACAGAGACTTCAATGACGCTCATCAACCCCAATCTCAACACAGATGATGATTC 2178
QY 819 CysGluArgValTyrPysArgGlyThrThrGluCysSerGlyCysPheArgSerArgSer 838
DB 2179 TGTGACAACTTTGGAAAGGACCCGAAACAA-----CGTCGGCAA 2220
QY 839 ArgGlyLysGlnAlaArgLysSer-----TrpAlaGlyLeuGlnLeuGlyLys 856
DB 2221 AGGCCAGACATACCCAGGAGACAGTACAAATGGAAGACCTTACAAAGAAATTAAGAA 2280
QY 857 MetLysSerArgValHisAlaGlnHisAsnArgThrProArgGluVal 872
DB 2281 GGTGAACACAGGAGACCAAGCAATTTGAGAGGGACCCAGAGAGTGTG 2328

```

RESULT 12
AA087442
ID AA087442 standard: cDNA; 2601 BP.

XX AA087442;
XX 21-NOV-1995 (first entry)
XX Human semaphorin III cDNA.
XX Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;
XX variola major virus; smallpox; semaphorin receptor binding activity;

```

KW modulation; nerve cell growth; immune response; viral pathogenesis;
KW neurological disease; neuro-regeneration; oncological infection; ds.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 16..2331
XX FT
XX FT /tag=a
XX FT /product= human semaphorin III
XX
XX W09507706-A.
XX
XX 23-MAR-1995.
XX
XX 13-SEP-1994; 94MO-US10151.
XX
XX 13-SEP-1993; 93US-0121713.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Bentley DR, Goodman CS, Kolodkin AL, Mattes D;
XX O'Connor T;
XX
XX WPI; 1995-131177/17.
XX
XX P-SDB; AAR71380.
XX
XX New class of semaphorin peptide(s) and polypeptide(s) - are
XX potent modulators of nerve cell growth and regeneration
XX
XX Example 2; Page 60-63; 101pp; English.
XX
XX The sequence of the cDNA encoding the human semaphorin III protein.
XX The proteins encoded by the grasshopper semaphorin I (AA087441), human
XX semaphorin III, vaccinia virus semaphorin IV (AA087443), human
XX semaphorin I and II (AA087444-5), Tribolium semaphorin I (AA087446) or
XX variola major (smallpox) virus semaphorin IV (AA087447) genes were used
XX to generate a series of peptides (AAR70370-R70418), which retain
XX semaphorin receptor binding activity. The semaphorin derived or
XX semaphorin receptor derived peptides are potent modulators of nerve cell
XX growth, immune responsiveness and viral pathogenesis. They can be used
XX in diagnosis and treatment of neurological disease and
XX neuro-regeneration, immune modulation and diagnosis and treatment of
XX viral and oncological infection and diseases.
XX
XX
XX Sequence 2601 BP; 809 A; 533 C; 593 G; 666 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 9, 04e-92 Length: 2601
XX Score: 1830.00 Matches: 369
XX Percent Similarity: 63.948 Conservative: 140
XX Best Local Similarity: 46.368 Mismatches: 246
XX Query Match: 38.568 Indels: 42
XX DB: Gaps: 13
XX
XX US-09-813-290-2 (1-875) x AA087442 (1-2601)
XX
XX 96 ProSerAlaTrpAla-----IleCysTrpLeuGlnGlyLeuLeuHis 111
XX |||:||||| ||| ||| |||:|||||
XX DB 8 CCGCAGACATGCGCTGTAACTAGATTTCTGT-CTTTTCGGGAGATTAATCTTACA 66
XX
XX 112 GlyGlySerSerGlyProSerProGlyProSerValProArgLeuArgLeuSerTyrArg 131
XX |||:||||| ||| ||| |||:|||||
XX DB 67 GCAGAGCAAACTATGAGATGGGAGAACAAATGTCGCAAGGCTGAATTAATCTTCAAA 126
XX
XX 132 AspLeuLeuSerAlaAsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeu 151
XX |||:||||| ||| ||| |||:|||||
XX DB 127 GAATGTTGGAATCCAAATGATGATGACTTTCAATAGGCTTGGCCAAACCTCCAGTTAT 186
XX
XX 152 GlnAlaMetCysTrpLeuAspGluTyrArgAspArgLeuPheLeuGlnGlyLeuAspAlaLeu 171
XX |||:||||| ||| ||| |||:|||||
XX DB 187 CATACCTTCTTTGGATGAGAACGAGAGAGCTGTATGTTGAGCAAGATCACATA 246
XX
XX 172 TyrSerLeuArgLeuAspGlnAlaTrpProAspProArgGluValLeuTrpProGln 191

```


XX AAX89112;
 AC
 XX 14-SEP-1999 (first entry)
 DT
 XX
 DE Human brain tissue-derived polypeptide coding sequence (clone OM007).
 XX
 XX Brain tissue; human; bone marrow; umbilical cord venous endothelial cell;
 KM recombinant; diagnosis; treatment; ss.
 XX
 OS Homo sapiens.
 XX MO9933873-A1.
 PN
 XX 08-JUL-1999.
 PD
 XX
 PF 25-DEC-1998; 98MO-JP05952.
 XX
 PR 26-DEC-1997; 97JP-0358811.
 XX
 PA (ONOY) ONO PHARM CO LTD.
 XX
 PI Fukushima D, Shibayama S, Tada H;
 XX
 DR WPI: 1999-419088/35.
 DR P-PSDB; AAY27127.
 XX
 PT New adult human brain tissue-produced polypeptides useful for
 PT diagnosis and treatment
 PS
 XX
 PS Claim 4; Page 39-40; 86pp: Japanese.
 XX
 CC The invention provides polypeptides (AAY27127-Y27133) produced by human
 CC adult brain tissue, human bone marrow or a human umbilical cord venous
 CC endothelial cell. Host cells transformed with vectors comprising the
 CC nucleic acids encoding the polypeptides are used for the recombinant
 CC expression of the polypeptides. The polypeptides can be used in
 CC diagnosis, treatment and basic studies. With wide applications in
 CC treatment depending on the activity to be aimed at. Sequences
 CC AAX89112-125 represent nucleic acids encoding the polypeptides.
 XX
 SQ Sequence 2331 BP; 723 A; 476 C; 529 G; 603 T; 0 other;
 Alignment Scores:
 Pred. No.: 7 77e-86 Length: 2331
 Score: 1721.00 Matches: 346
 Percent Similarity: 63.89% Conservative: 144
 Best Local Similarity: 45.11% Mismatches: 229
 Query Match: 36.26% Indels: 48
 DB: 20 Gaps: 17
 US-09-813-290-2 (1-875) x AAX89112 (1-2331)
 QY 122 SerValProArgLeuArgLeuSerTyrArgAspLeuLeuSerAlaAsnArgSerAlaIle 141
 DB 121 AATATTCACAACTCAAGCTACCTCAACAAAGACTCTGCTTCAATAGCTATATCC 180
 QY 142 PheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTyrLeuAspGlyTyrArgAsp 161
 DB 181 TTTTGGGTTTCACAGACGAGCTGGATTTTCAAACTTCTCTTAATGAGGAAAGAGCC 240
 QY 162 ArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeuAspGlnAlaIleTyrPro 181
 DB 241 AGGCGTCTTGGGAGCCAAAGACACATCTTCTACAGTCTGCTTACCTTAACAA 300
 QY 182 AspProArgGluValLeuTyrProGlnProGlnArgGlnArgGlnGlyValArgGly 201
 DB 301 AATTTTAAGAAATTTATTTGGCTCTGCAAGAAAGAGGCTGAATTAATTAATAGCT 360
 QY 202 GlyArgAspProLeuThrGluCysAlaAsnPheValArgValLeuGlnProHisAsnArg 221
 DB 361 GGGAAAGATGCCAATACAGAAATGTCAATTTCTACAGAGTACTTACGCCCTATAACAA 420

QY 222 ThrHisLeuAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeuIleThrVal 241
 DB 421 ACTCATAATATATGTGTGGAACTGGACATTTCAATATGTGGATATATATCTT 480
 QY 242 Gly---HisArgGlyGlnHisValLeuHisLeuGlnProGlySerValGluSerGlyArg 260
 DB 481 GGAGTCTACAAAGGAGATATTTATTTCAAACTAGACACACATTAATTTGGAGTCTGGCAGA 540
 QY 261 GlyArgCysProHisGluProSerArgProPheAlaSerThrPheIleAspGlyGluLeu 280
 DB 541 CTGAATCTCTTTCAGATCTGACAGCCTTTTGGTTAGTAATGACAGATGAGAGCTC 600
 QY 281 TyrThrGlyLeuThrAlaAspPheLeuGlyArgGlyAlaMetIlePheArgSerGlyGly 300
 DB 601 TACTCTGGACAGCTTCTGATTTCTTGGCAAGATCTGATTCACATTCAGATCTTGGG 660
 QY 301 Pro-----ArgProAlaLeuArgSerAsp---SerAspGlnSerLeuLeuHisAsp 316
 DB 661 CCTACTCATGACCCACACCTACATCGAATCTACATTTTCAGACACCTAGCTGCTCAATGGA 720
 QY 317 ProArgPheValMetAlaAlaArgIleProGluAsnSerAspGlnAspAsnAspVal 336
 DB 721 GCAAATTTTATTTGGAACTTCTTCATACACACACACATTCACATTCAGATGATGAATA 780
 QY 337 TyrPhePheSerGlyThrValProSerProAspGlyGlySerAsnHisValThrVal 356
 DB 781 TATTTCTTCTTCTGCTGAATCA-----TCTCAAGAGACAGCTACCTCGATTAACCATC 834
 QY 357 ---SerArgValGlyArgValCysValAsnAspAlaGlyGlyGlnArgValLeuValAsn 375
 DB 835 CTTCCTCGAGTTGGAGAGGTTTGAAGATGATGTAGAGAGACACACACAGCTGATTAAC 894
 QY 376 LysThrSerThrPheLeuValArgLeuValCysSerValProGlyProGlyGlyAla 395
 DB 895 AAGGAGCGACTTTTCTTAAGCCACAGCTATTTCTCAATCTCGAAGTGAAGTGAAGGCA 954
 QY 396 GluThrHisPheAspGlnLeuGluAspValPheLeuLeuTyrProLysAlaGlyLysSer 415
 DB 955 GATCTACTTATTTGATGAGCTTCAAGATATTTATTTACTCCACACAAAGATGAAGAAT 1014
 QY 416 LeuGluValTyrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCys 435
 DB 1015 CTTGTATATATGAGCTTCTTACTACACACACCTCTTCAAGAGCTGCTGTTGT 1074
 QY 436 ValTyrHisMetAlaAspIleTyrGluValPheAsnGlyProPheAlaHisArgAspGly 455
 DB 1075 GTGTATAGCATGGCTGACATCAGACAGCTTTTATATGCTCCATATATGCTCAAGGAAGT 1134
 QY 456 ProGlnHisGlnTyrGlyProTyrGlyGlyLysValProPheProArgProGlyValCys 475
 DB 1135 GCAGACCATGCTGGGCGAGTATGATGAGGAGAATTCCTATTCACAGGCTGTACATGT 1194
 QY 476 ProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTyrProAsp 495
 DB 1195 CCAAGCAAA---ACSTATGACCA-----CTGATTAAAGTCCACCGAGATTTTCCAGAT 1245
 QY 496 GluValLeuGlnPheAlaArgAlaHisProLeuMetPheThrProValArgProGlnHis 515
 DB 1246 GATGTATATGATGTTTCAATAAGCGGCACTGTGTATATATAGTCCGTATACCAAGTTGCA 1305
 QY 516 GlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleValAsp 535
 DB 1306 GGGAGCAACAGCTTCAAGAGAAATCAATGTGATTTACAGATCACAGATAGTGGTGAAT 1365
 QY 536 ArgValGluAlaGluAspGlyThrTyrAspValIlePheLeuGlyThrAspSerGlySer 555
 DB 1366 CATGTATTTGCAAGATATGCGCAGTACGATGATATTTCTTGGAGACACATGTAAGTGAAT 1425
 QY 556 ValLeuLysValIleAlaLeuGlnAlaGlyGlySerAlaGluPro-----GluGlu 572
 DB 1426 GTTCCTCAAAAGTTGTGACGAT-----TCAAGAGAAAGTGAATATGGAAGG 1473
 QY 573 ValValLeuGlnGluLeuGlnValPheLysValProThrProIleThrGluMetGluIle 592

```

Db      1474  GTAGTCTGAGGAGATTGCAATATTCACAGCACTCATCATATCTTGAACATGGAATG 1533
OY      593  SerValIysArgIleMetLeuTyrValIglYserArgLeuGlyValAlaGlnLeuArgLeu 612
OY      1534  TCTCTGAGGAGCAACAATTTGATGCTGGTCCGAGATGATTAGTTCAGCTCTCTTG 1593
OY      613  HIsGlnCysGluThrTyrGlyThrAlaCysAlaGluCysCysLeuAlaArgAspProTyr 632
OY      1594  CACAGATGCACACTATATGGCAAGCTTGGCAGACCTGTGTCTTCCAGACAGCCCTTAC 1653
OY      633  CysAlaThrAspGlyAlaSerCysThrHisTyrArgProSerLeuGlyLysArgArgPhe 652
OY      1654  TGTCTCTGGGATGGAATGCAATGCTCTCGATATGCTCTACT--TCTAAAGAGAGCT 1710
OY      653  ArgArgGlnAspIleArgHisGlyAsnProAlaLeuGlnCys----- 666
OY      1711  AGACGCCAAGATGTAAATATGAGGACCAATCACCCAGTCTGGACATCGAAGACAGC 1770
OY      667  LeuGlyGlnSerGlnGluGluAlaValGlyLeuValAlaAlaThrMetValTyrGly 686
OY      1771  ATTACTCATGAACCTGCTGATGAAAGGTG-----ATTTTGGC 1809
OY      687  ThrGlnHisAsnSerThrPheLeuGlnCysLeuProLysSerProGlnAlaAlaValArg 706
OY      1810  ATTGAATTTAACTCAACCTTCTGTGAATGTATACCTAAATCCCAACAGCACTATTAA 1869
OY      707  TyrLeuLeuGlnArgProGlyValAspGlyProAspGlnValLysThrAspGluArgVal 726
OY      1870  TGGTATATCCAGAGGTCAAGGATGACATGACATGAGAGATTGAACCGATGAAGAATC 1929
OY      727  LeuHisThrGluArgGlyLeuLeuPheArgArgLeuSerArgPheAspAlaGlyThrTyr 746
OY      1930  ATCAAAACGGAATATGGCTACTGATTCAGAGTTTCAGACAGAACAGAGATCTGGATGAT 1989
OY      747  ThrCysThrThrLeuGlnHisGlyPheSerGlnThrValAlaArgLeuAlaLeuValVal 766
OY      1990  TACTGAAAGCCAGAGACACACTTTCATCCACACCATGATGAGCTGACTTGAATGTC 2049
OY      767  IleValAlaSerGlnLeuAsnLeuAsnLeuPheProProGlnProLysProGlnGluProPro 786
OY      2050  ATTGAAGATTAACAGATGAGAAAT-----ACCCAGAGGCGAGAGATAGAGGGGCGAG 2103
OY      787  AlaArgGlyLysLeuAlaSerThrProProLysAlaThrTyrLysAspIleLeuGlnLeu 806
OY      2104  GTCAAGGATTAATGGCTGAGTCA-----CGGTGAGATACAAAGACTATCAAAATC 2157
OY      807  IleGlyPheAlaAsnLeuProArgValAspGluTyrCysGluArgValITPCysArgGly 826
OY      2158  CTTAGCAGCCCAAACTTC--AGCCTGACCACTACTGCGAACACGATGTG----- 2205
OY      827  ThrThrGluCysSerGlyCysPheArgSerArgSerArgGlyLysGlnAlaArgGlyLys 846
OY      2206  -----CACAGGAGAGAGCGGAGACGAGAAACAGAGGGGGGCCA 2244
OY      847  SerTrpAlaGlyLeu---GluLeuGlyLysLysMetLysSerArgValHisAlaGlnHis 865
OY      2245  AAGTGAAGCAACATGCAGAAATGAGAAAGAAATCGAAGACATCACAGACCTG 2304
OY      866  AsnArgThrProArgGluVal 872
OY      2305  GATGAGCTCCCTAGAGCTGTA 2325
Db

```

RESULT 14
 AAA37109
 ID AAA37109 standard: cDNA, 3871 BP.
 AC AAA37109;
 XX
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1491 (UNC760) cDNA sequence SEQ ID NO:309.
 XX

```

KW      Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW      transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW      ss.
XX      Homo sapiens.
OY      WO200012708-A2.
OY      09-MAR-2000.
OY      01-SEP-1999; 99WO-US20111.
OY      01-SEP-1998; 98US-0098716.
OY      01-SEP-1998; 98US-0098749.
OY      01-SEP-1998; 98US-0098750.
OY      02-SEP-1998; 98US-0098803.
OY      02-SEP-1998; 98US-0098821.
OY      02-SEP-1998; 98US-0098843.
OY      02-SEP-1998; 98US-0099536.
OY      09-SEP-1998; 98US-0099596.
OY      09-SEP-1998; 98US-0099598.
OY      09-SEP-1998; 98US-0099602.
OY      09-SEP-1998; 98US-0099642.
OY      09-SEP-1998; 98US-0099741.
OY      10-SEP-1998; 98US-0099754.
OY      10-SEP-1998; 98US-0099792.
OY      10-SEP-1998; 98US-0099792.
OY      10-SEP-1998; 98US-0099808.
OY      10-SEP-1998; 98US-0099812.
OY      10-SEP-1998; 98US-0099815.
OY      10-SEP-1998; 98US-0099815.
OY      10-SEP-1998; 98US-0100385.
OY      15-SEP-1998; 98US-0100388.
OY      15-SEP-1998; 98US-0100390.
OY      15-SEP-1998; 98US-0100584.
OY      16-SEP-1998; 98US-0100627.
OY      16-SEP-1998; 98US-0100661.
OY      16-SEP-1998; 98US-0100662.
OY      16-SEP-1998; 98US-0100664.
OY      17-SEP-1998; 98US-0100683.
OY      17-SEP-1998; 98US-0100684.
OY      17-SEP-1998; 98US-0100710.
OY      17-SEP-1998; 98US-0100711.
OY      17-SEP-1998; 98US-0100919.
OY      17-SEP-1998; 98US-0100930.
OY      17-SEP-1998; 98US-0100849.
OY      18-SEP-1998; 98US-0101014.
OY      18-SEP-1998; 98US-0101068.
OY      18-SEP-1998; 98US-0101071.
OY      22-SEP-1998; 98US-0101279.
OY      23-SEP-1998; 98US-0101471.
OY      23-SEP-1998; 98US-0101472.
OY      23-SEP-1998; 98US-0101474.
OY      23-SEP-1998; 98US-0101475.
OY      23-SEP-1998; 98US-0101476.
OY      23-SEP-1998; 98US-0101477.
OY      23-SEP-1998; 98US-0101479.
OY      23-SEP-1998; 98US-0101738.
OY      24-SEP-1998; 98US-0101741.
OY      24-SEP-1998; 98US-0101743.
OY      24-SEP-1998; 98US-0101915.
OY      24-SEP-1998; 98US-0101916.
OY      29-SEP-1998; 98US-0102207.
OY      29-SEP-1998; 98US-0102240.
OY      29-SEP-1998; 98US-0102307.
OY      29-SEP-1998; 98US-0102330.
OY      29-SEP-1998; 98US-0102331.
OY      30-SEP-1998; 98US-0102484.
OY      30-SEP-1998; 98US-0102487.
OY      30-SEP-1998; 98US-0102570.
OY      30-SEP-1998; 98US-0102571.
OY      01-OCT-1998; 98US-0102684.

```



```

Db 941 CTTTCTCGAGTTGGAGAGTTTGTAGAAATGATGAGAGACAGCACCGCTGATTAAC 1000
Qy 376 LysTPSerThrpheLeuLysAlaGleuValCysSerValProGlyProGlyGlyAla 395
Db 1001 AAGTGGACGACTTTCTTAAGGCCAGACTGATTGTGCATTTCTGGAGAGTATGGGCA 1060
Qy 396 GluThriHisPheAspGlnLeuGluAspValPheLeuLeuTrpProLysAlaGlyLysSer 475
Db 1061 GATTCCTACTTGTGAGCTTCAAGATATTATTACTCCCAACAGAGATGAAGAAAT 1120
Qy 416 LeuGluValTyralaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCys 435
Db 1121 CCGTACTAATATGAGACTTACTACCAACAGCTCCATCTTCAAGGCTGCTGCTTTGT 1180
Qy 436 ValTyriHisMetAlaAspLiletrpGluValPheAsnGlyProPheAlaHisArgAspGly 455
Db 1181 GTGTATGACAGGCTGATCAGACAGCACTTTTATGAGTCCATATGCTCATTAAGAAAGT 1240
Qy 456 ProGlnHisGlnTrpGlyProTyriGlyLysValProPheProAlaGProGlyValCys 475
Db 1241 GCAGACCACTGCTGGGTCAGATATGATGAGAAATTCCTATCCAGCCCTGCTACATGT 1300
Qy 476 ProSerLysMetThrAlaGlnProGlyLarProPheGlySerThrLysAspTrpProAsp 495
Db 1301 CCAAGCAAA--ACCTATGACCA--CTGATTAAAGTCCACCCGAGATTTCACAGAT 1361
Qy 496 GluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpProValAlaArgProAlaHis 515
Db 1362 GATGTCATCAGTTTCATTAAGGGGCACTCTGATGTATTAAGTCCGTATACCCAGTTGCA 1411
Qy 516 GlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnLleValValAsp 535
Db 1412 GGAGGACCAACGCTTCAGAGCAATGATGATTCAGAGCTACACAGATAGTGGGAT 1471
Qy 536 ArgValGluAlaGluAspLysThrTyriAspValLlePheLeuGlyThrAspSerGlySer 555
Db 1472 CATGTCATTCAGAGAGATGCGCAGTACGATGTAATGTTCTTGAGACAGACATGGAGACT 1531
Qy 556 ValLeuLysValLleAlaLeuGlnAlaGlyGlySerAlaGluPro-----GluGlu 572
Db 1532 GTCCTCAAGTGTGTGACAT-----TCAAGAGAAAGTGAATATGAGAGAG 1579
Qy 573 ValValLeuGlnGluLeuGlnValPheLysValProThrProLleThrGlnMetGluLe 592
Db 1580 GATGCTCTGAGGAGTTCGATATTCAGACACTCATCATCTTGAACATGAATG 1639
Qy 593 SerValLysArgGlnMetLeuTyriValGlySerArgLeuGlyValAlaGlnLeuArgLeu 612
Db 1640 TCTCTGAGCAGCAACAATGTGATGTTCCCGAGATGATTAAGTTCAGCTCTCTGTG 1699
Qy 613 HisGlnCysGluThrTyriGlyThrAlaCysAlaGluCysCysLeuAlaArgAspProTyri 632
Db 1700 CACAGATGCGCACCTTATGGAAGCTTGGCAGACTGTTGCTCTGGCAGAGACCCCTAC 1759
Qy 633 CysAlaLarPaspGlyAlaSerCysThrHisTyriArgProSerLeuGlyLysAlaArgPhe 652
Db 1760 TGTGCTGGAGTGAATATGATGCTCTCGATATGCTCTACT--TCTAAAGAGAGACT 1816
Qy 653 ArgArgGlnAspLileArgHisGlyAspProAlaLeuGlnCys-----666
Db 1817 AGACGCCAAGGTGTAATATATGAGGCCCAATCCAGTCTGGGACATCGAAGACAGC 1876
Qy 667 LeuGlyGlnSerGlnGluGluAlaValAlaGlyLeuValAlaAlaThrMetValTyriGly 686
Db 1877 ATTAGTCAATGAACGTCGATGAAGAGTG-----ATTTTGGC 1915
Qy 687 ThrGluHisAsnSerThrPheLeuGluCysLeuProLysSerProGlnAlaAlaValArg 706
Db 1916 ATTGAATTTAATCAACCTTCTGGAATGTATACCTAATCAACAGCAACTTTTAA 1975
Qy 707 TrpLeuGlnArgProGlyAspGlnGlyProAspGlnValLysThrAspGluValArg 726

```

```

Db 1976 TGTATATCCAGAGCTCAGGCGATGACATCGAGAGAGTGTGAAGCCCGCATGAAGATC 2035
Qy 727 LeuHisThrGluArgGlyLeuLeuPheAlaArgLeuSerArgPheAspAlaGlyThrTyri 746
Db 2036 ATCAAAACGGAAATATGGGCTACTGATTCGAGATTTCAGAGAAAGAGATTCTGGGATGTAT 2095
Qy 747 ThrCysThrThrLeuGlnHisGlyPheSerGlnThrValValAlaArgLeuAlaLeuValAl 766
Db 2096 TACTGCAAAAGCCCGAGAGCAGACACTTTCATCCACACANAGTGAAGCTGACTTGAATGTC 2155
Qy 767 LLeValAlaSerGlnLeuAspAsnLeuPheProProGluProLysProGluGluProPro 786
Db 2156 ATTGAGATGACAGATGGAAGAT-----ACCGAGAGCGCAGCATGAGAGGGCGAG 2209
Qy 787 AlaArgGlyGlyLeuAlaSerThrProProLysAlaTrpTyriLysAspLleLeuGlnLeu 806
Db 2210 GTCAGAGATCTATTTGGCTGACTG-----CGGTGACATACAAAGACTCATCAATTC 2263
Qy 807 LLeGlyPheAlaAsnLeuProAlaArgValAspGluTyriCysGluArgValLTrpCysArgGly 826
Db 2264 CTTACAGCCCAACTTC--AGCTCGACACAGTACTCGAAGACAGATGTGG-----2311
Qy 827 ThrThrGluCysSerGlyCysPheArgSerArgSerArgGlyLysGlnAlaArgGlyLys 846
Db 2312 -----CACAGGAGAAAGCGAGACAGAGAAACAAAGGGGGCCCA 2350
Qy 847 SerTrpAlaGlyLeu---GluLeuGlyLysLysMetLysSerArgValHisAlaGlnHis 865
Db 2351 AACTGGAACCATGCAATGCAAGAAATGAAAGAAATCGAAGACATCAGACAGACTG 2410
Qy 866 AsnAlaGlyThrProArgGluVal 872
Db 2411 GATGACCTCCTAGACTGTGA 2431
RESULT 15
ID AAS46098 standard: cDNA; 3871 BP.
XX AAS46098;
XX
XX 18-DEC-2001 (first entry)
DE Human DNA encoding PRO polypeptide sequence #174.
XX
XX PRO polypeptide: mammal; tumour; cancer; human; cattle; horse; sheep; ss;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
XX PCR primer.
XX
XX Homo sapiens.
XX
XX WO200168848-A2.
XX
XX 20-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US06520.
XX
XX 01-MAR-2000; 2000WO-US05601.
XX
XX 02-MAR-2000; 2000WO-US05841.
XX
XX 03-MAR-2000; 2000US-187202P.
XX
XX 06-MAR-2000; 2000US-186968P.
XX
XX 14-MAR-2000; 2000US-189320P.
XX
XX 14-MAR-2000; 2000US-189328P.
XX
XX 15-MAR-2000; 2000WO-US06884.
XX
XX 21-MAR-2000; 2000US-190828P.
XX
XX 21-MAR-2000; 2000US-191007P.
XX
XX 21-MAR-2000; 2000US-191048P.
XX
XX 21-MAR-2000; 2000US-191314P.
XX
XX 28-MAR-2000; 2000US-192655P.
XX
XX 29-MAR-2000; 2000US-193032P.
XX
XX 29-MAR-2000; 2000US-193053P.
XX
XX 30-MAR-2000; 2000WO-US08439.
XX

```

PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 03-MAY-2000; 2000US-199654P.
 PR 17-MAY-2000; 2000US-201516P.
 PR 22-MAY-2000; 2000US-201370S.
 PR 30-MAY-2000; 2000US-201404Z.
 PR 02-JUN-2000; 2000US-2015264.
 PR 28-JUL-2000; 2000US-209832P.
 PR 22-AUG-2000; 2000US-2020710.
 PR 24-AUG-2000; 2000US-0644848.
 PR 08-NOV-2000; 2000US-0523328.
 PR 01-DEC-2000; 2000US-0530952.
 PR 20-DEC-2000; 2000US-0532678.
 PR 20-DEC-2000; 2000US-0534956.

(GETH) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL,
 Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

WPI: 2001-602746/68.

P-PSDB: AAU29197.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 presence of tumours, such as prostate and breast tumours, in mammals and
 to screen for modulators of the compounds -

Claim 2; Fig 347; 774p; English.

Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
 primers for PRO polypeptides of the invention. The sequences of the
 invention can be used to detect the presence of a tumour in a mammal by
 comparing the level of expression of a PRO polypeptide in a test sample
 of cells from the animal and a control sample of normal cells, whereby a
 higher level of expression in the test sample indicates the presence of a
 tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 pigs, goats and rabbits but are preferably human. The polypeptides can be
 used to stimulate tumour necrosis factor (TNF) alpha release from human
 blood, when contacted with it. A specific polypeptide can be used to
 stimulate the proliferation or differentiation of chondrocyte cells. The
 PRO proteins can be used to determine the presence of tumours and also
 susceptibility to tumour development, particularly adrenal, lung, colon,
 breast, prostate, rectal, cervical, or liver tumours, in mammalian
 subjects. The oligonucleotide probes specific for the PRO nucleic acids
 can be used for genetic analysis of individuals with genetic disorders.

Sequence 3871 BP; 1251 A; 704 C; 805 G; 1111 T; 0 other;

Alignment Scores:

Pred. No.: 1.33e-85 Length: 3871
 Score: 1721.00 Matches: 346
 Percent Similarity: 63.89 Conservative: 144
 Best Local Similarity: 45.118 Mismatches: 229
 Query Match: 36.268 Indels: 48
 DB: 22 Gaps: 17

US-09-813-290-2 (1-875) x AAS46098 (1-3871)

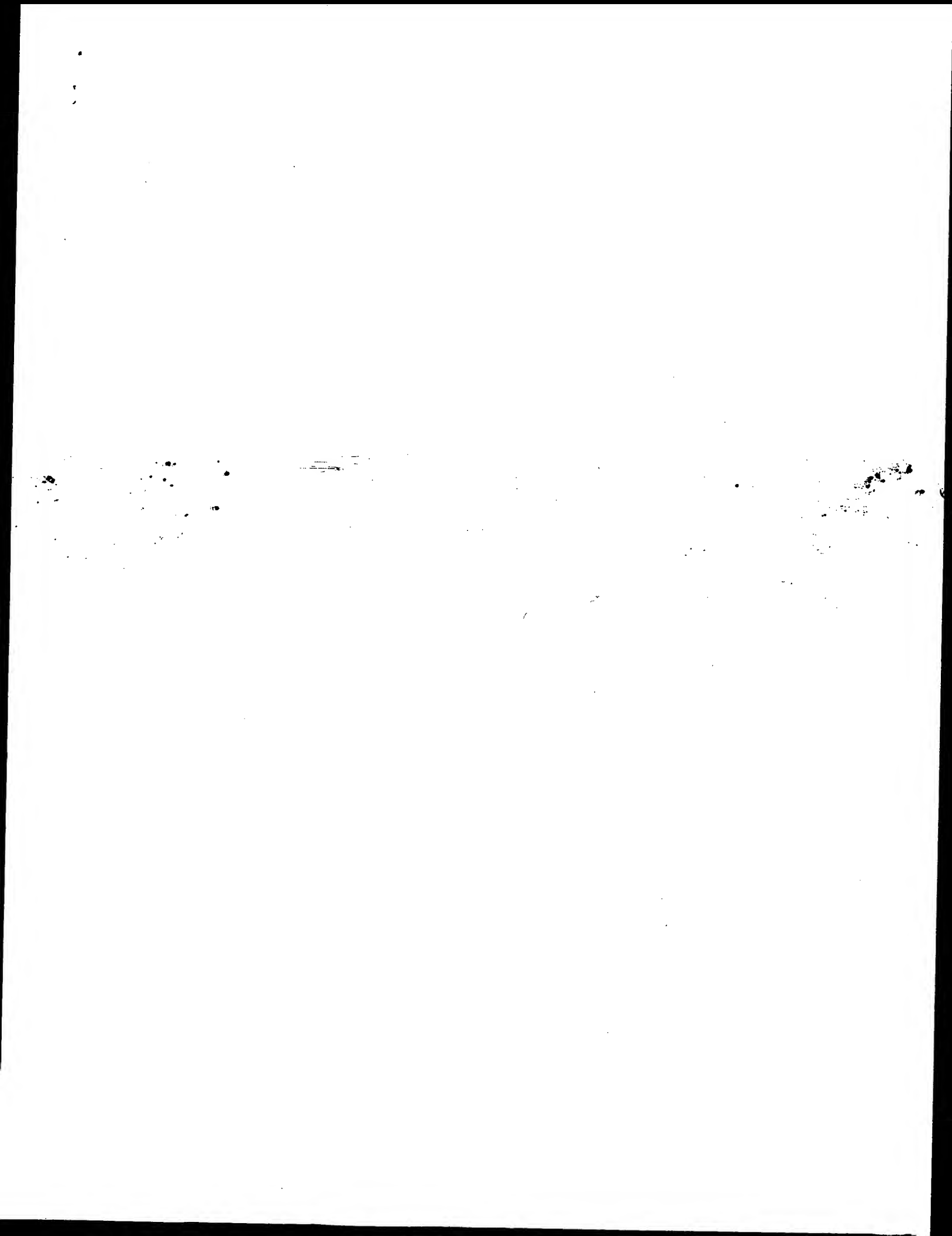
QY 122 SerValProAlaLeuArgLeuSerTyrArgAspLeuLeuSerAlaAsnArgSerAlaIle 141
 Db 227 AATATTCACAGACTCAAGCTCAAGCTCAAGCTCTTCAATAGCTGATATCC 286
 QY 142 PheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTyrLeuAspGlnTyrArgAsp 161

Db 287 TTTTGGTTCATCAGAGAGACTGATTTCAACTCTTCTAGATGAGGAAGAGGC 346
 QY 162 ArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeuAspGlnAlaTyrPro 181
 Db 347 AGCTGCTCTGGAGAGCCAAAGACACACATCTTCTACTCAGTCTGGTGAATTAAACAA 406
 QY 182 AspProArgGlnValLeuTyrProGlnProGlyValInArgGlyGlyValInGly 201
 Db 407 AATTTTAAAGAAAGTTTATTTGCTGCTGCAAGAGAGAGGGGTGAATTAATTAATGCT 466
 QY 202 GlyArgAspProLeuThrGlyCysAlaAsnPheValArgValLeuGlnProHisAsnArg 221
 Db 467 GGGAAAGATGCCAATACAGATGTGCAAAATTTATCATCAGACTTCAAGCCCTTAACAA 526
 QY 222 ThrHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeuIleThrVal 241
 Db 527 ACTCACATATATGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 586
 QY 242 Gly---HisArgGlyGlyLeuHisValLeuHisLeuGlnProGlySerValGlySerGlyArg 260
 Db 587 GGAGTCTACAGAGGAGATTTATATCAACATGACACACATATTTGGAGTGTGAGTGTGAGTGT 646
 QY 261 GlyArgCysProHisGlnProSerArgProPheAlaSerThrPheIleAspGlyLeu 280
 Db 647 CTGAATGTCTCTTTCGATCTCAGACGCTTGTGCTGAGTATGACATGACATGATGATGATGAT 706
 QY 281 TyrThrGlyLeuThrAlaAspPheLeuGlyArgGlnAlaMetIlePheArgSerGlyGly 300
 Db 707 TACTGTGAAACAGCTTGTGATTTCTTGGCAAAAGATGATGATGATGATGATGATGATGAT 766
 QY 301 Pro-----ArgProAlaLeuArgSerAsp---SerAspGlnSerLeuLeuHisAsp 316
 Db 767 CCTACTCATGACACACACATCAATCAAGATGATGATGATGATGATGATGATGATGATGATGAT 826
 QY 317 ProArgPheValMetAlaAlaArgIleProGlnAsnSerAspGlnAspAsnArgVal 336
 Db 827 GCMAAATTTATGGAACCTTCTTACATACAGACACCTCAATCCAGATGATGATGATGATGAT 886
 QY 337 TyrPhePhePheSerGlnThrValProSerProAspGlyGlySerAsnHisValThrVal 356
 Db 887 TATTTCTTCTTTCGGAATCA-----TCTCAGAGAGCAGTACTCCGATTAACCAATC 940
 QY 357 ---SerArgValGlyArgValCysValAsnAspAlaGlyGlyGlnArgValLeuValAsn 375
 Db 941 CTTCCTCAGATTTGGAAGCTTTGTAAGATATGATGATGATGATGATGATGATGATGATGAT 1000
 QY 376 LysTrpSerThrPheLeuLeuAlaArgLeuValCysSerValProGlyGlyVal 395
 Db 1001 AAGTGGAGACTTTTCTTAAGCCAGACTGATTTGCTCAATTCCTGGAAGTGAAGGGGCA 1060
 QY 396 GlnThrHisPheAspGlnLeuGlnAspValPheLeuLeuTyrProGlyAlaGlySer 415
 Db 1061 GATACCTACTTGTGATGACCTTCAAGATATTTATTTACTCCCAACAGAGTGAAGAAAT 1120
 QY 416 LeuGlnValTyrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCys 435
 Db 1121 CCGTGTATATATGAGCTTCTTACTCAACACCTCAATCTTCAAGGCTCGCGTTGT 1180
 QY 436 ValTyrHisMetAlaAspIleTyrPheGlnValPheAsnGlyProPheAlaHisArgAspGly 455
 Db 1181 GTGTATGACATGCTGATCAGACAGCAGTTTAAATGCTCAATATGCTCAATAGGAAGT 1240
 QY 456 ProGlnHisGlnTyrPheGlyGlyValPheValProPheProArgProGlyValCys 475
 Db 1241 GCAGACCATGTTGGGCGAGTATGATGGAGAAATTCCTATATCAGCGCTGTACTGT 1300
 QY 476 ProSerTyrMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTyrProAsp 495
 Db 1301 CCAAGCAAA---ACCTATGACCA-----CTGATTAAAGTCCACCGAGATTTTCCAGAT 1351
 QY 496 GlnValLeuGlnPheAlaArgAlaHisProLeuMetPheThrProValArgProAspHis 515

Db 1352 GATGTCATCATGTTTCATTAAGCGCACTGTGATGATATAGTCCTATACCCAGTTGCA 1411
OY 516 G1YArgProValLeuValIysThrHisLeuAlaGlnGlnLeuHisGlnLeuValAla 635
1412 GGAGGACCAACGTTCAAGAGATCAATGATGATTACAGCTGACACGATGATGGTGAT 1471
OY 536 ArgValGlnIleuIleuAspGlyThrTyrAspValIlePheLeuGlyThrAspSerGlySer 555
1472 CATGTCATTCAGAGAGTGGCCAGTACGATGATGATGTTCTTGGAAACAGACATGGAAT 1531
OY 556 ValLeuIysValIleAlaLeuGlnAlaGlyIleSerAlaGluPro-----GluGlu 572
1532 GTCCTCAAGATTGTCACACT-----TCAAGAGAAAGTGAATATGGAAG.1579
OY 573 ValValLeuGlnGlnLeuGlnValPheIysValProThrProIleThrGlnLeuIle 592
1580 GTAGTGTGGAGAGTTCAGATTCACAGCATCATCATCATCTTGAACATGGAATG 1639
OY 593 SerValIysArgGlnMetLeuTyrValGlySerArgLeuGlyValAlaGlnLeuArgLeu 612
1640 TCTCTGAAGCAGCAACATGTTACATGTTCCGAGATGATGATTCACCTCTCTG 1699
OY 613 HisGlnCysGlnThrTyrGlyThrAlaCysAlaGlnCysGlyLeuAlaArgAspProTyr 632
1700 CACAGATCGCACCTTATGAGAAAGCTTGGCAGACTGTTGTTGCCAGAGACCCCTAC 1759
OY 633 CysAlaTyrAspGlyAlaSerCysThrHisTyrArgProSerLeuGlyLysArgArgPhe 652
1760 TGTGCTGGAGTGAATGCAATGCTGCTGATGCTGCTACT--TCTAAAGCAGAGCT 1816
OY 653 ArgArgGlnAspIleArgHisGlnGlyAsnProAlaLeuGlnCys----- 666
1817 AGACCCCAAGATTAATAATGCGACCCCAATCCACCAGTGTGGAGACTCGAAGACAC 1876
OY 667 LeuGlyGlnSerGlnGlnGlnIleValGlyLeuValAlaIleThrMetValTyrGly 686
1877 ATTACTCATGAACCTGCTGATGAAGAGTG-----ATTTTGGC 1915
OY 687 ThrGlnHisAsnSerThrPheLeuGlnCysLeuProIysSerProGlnAlaValArg 706
1916 ATTGAATTAACTCAACTTCTGGAATGATATACCTAAATCCCAACACCACTATTAA 1975
OY 707 TyrPheLeuGlnArgProGlnIysAspGlnGlyProAspGlnValIysThrAspGlnArgVal 726
1976 TGTATATCCAGAGTGCAGGAGTGCATCGAGAGAGTGAAGCCGATGAAGAT 2035
OY 727 LeuHisThrGlnArgGlyLeuLeuPheArgArgLeuSerArgPheAspAlaGlyThrTyr 746
2036 ATCAAAACGGAATATGGCTACTGATTCGAACTTTCAGAAAGAGATTCCTGGAGTAT 2095
OY 747 ThrCysThrThrLeuGlnHisGlyPheSerGlnThrValAlaArgLeuAlaValVal 766
2096 TACTGCAAAACCCAGAGAGCAGCTTTCATCCACACCATGTAAGCTTGAATCTC 2155
OY 767 IleValAlaSerGlnLeuAspAsnLeuPheProGlnIleProIysProGlnIleProPro 786
2156 ATTGAGAAATGACAGATGAAT-----ACCAGAGGCGAGCATGAGAGGCGCAG 2209
OY 787 AlaArgGlyIleLeuAlaSerThrProIysAlaTyrIleValAspIleLeuGlnLeu 806
2210 GTCAAAGATCTATTGGCTGATCA-----CGTTGAGATACAAAGACTACATCCAAATC 2263
OY 807 IleGlyPheAlaAsnLeuProArgValAspGlnTyrCysGlnArgValTyrCysArgGly 826
2264 CTTAGCAGCCCAACTTC--AGCCTGACACAGTACGCAACAGATGTG----- 2311
OY 827 ThrThrGlnCysSerGlyCysPheArgSerArgSerArgGlyLysGlnAlaArgGlyLys 846
2312 -----CACAGGAGAGCGGAGAGACAGAGAACAGAGGCGCCEA.2350
OY 847 SerTyrPalaGlyLeu-----GluLeuGlyLysLysMetLysSerArgValHisAlaGlnHis 865
2351 AAGTGAAGCAGATGCAAGAAATGAGAAACGAATCGAAGACATCACAGACCTG 2410

OY 866 AsnArgThrProArgGlnVal 872
Db 2411 GATGAGCTCCCTAGAGCTGTA 2431

Search completed: October 9, 2002, 20:19:10
Job time : 348 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 14:12:18 ; Search time 44 Seconds
(without alignments)
1910.867 Million cell updates/sec

Title: US-09-813-290-2

Sequence: 1 MACALGKGVFPMSGWPVWHR.....KMSRVHAENRRPREVEAT 875

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*

1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1862	39.2	772	2	A49069
2	1832	38.6	772	2	I48747
3	1830	38.6	771	2	D49423
4	1747	36.8	749	2	G01856
5	1698.5	35.8	666	2	I58169
6	1658	34.9	753	2	G02173
7	1653.5	34.8	748	2	I48744
8	1568	33.0	751	2	I48748
9	862	18.2	834	2	S66498
10	831.5	17.5	782	2	I48746
11	676.5	14.3	760	2	I48745
12	628	13.2	730	2	JH0798
13	606	12.8	724	2	C49423
14	601.5	12.7	656	2	B49423
15	589	12.6	712	2	T27165
16	540.5	11.4	711	2	A49423
17	539.5	11.4	1074	2	JCS928
18	441.5	9.3	653	2	T03102
19	321.5	6.8	766	2	JCS3853
20	217.5	4.6	1894	2	JCS980
21	190	4.0	1945	2	T13937
22	187.5	4.0	1884	2	JCS975
23	185.5	3.9	1905	2	T5153
24	170	3.6	1568	2	T09074
25	162	3.4	403	2	E42521
26	162	3.4	2051	2	T13164
27	155	3.3	441	2	S29921
28	138.5	2.9	1872	2	JCS976
29	133	2.8	3375	2	T19621

30	132.5	2.8	295	2	J01775	Sal19R protein - v
31	117.5	2.5	1033	2	S19247	cell adhesion prot
32	116	2.4	686	2	E75267	prolyl endopeptida
33	116	2.4	1198	2	T28678	polyketide synthas
34	115.5	2.4	2082	2	T37056	probable multi-dom
35	113.5	2.4	2647	2	A37098	gelatin factor AB
36	113	2.4	952	2	S32954	hypothetical prote
37	112.5	2.4	620	2	T30765	hypothetical prote
38	111.5	2.3	6642	2	T29757	protein UNC-89 - C
39	111	2.3	678	2	C68495	hypothetical prote
40	111	2.3	678	2	H72128	3-methyl-2-oxobuta
41	111	2.3	1330	2	S49010	embryonic receptor
42	111	2.3	5175	2	T20992	hypothetical prote
43	111	2.3	5198	2	T43290	hemiscitin precurs
44	107.5	2.3	1276	2	T18526	SREBP cleavage act
45	107.5	2.3	2611	2	T14591	actinomycin synthe

ALIGNMENTS

RESULT 1

A49069
collapsin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Sep-1999
C:Accession: A49069
R:Uno, Y.; Raible, D.; Raper, J.A.
Cell 75, 217-227, 1993

A:Title: Collapsin: a protein in brain that induces the collapse and paralysis of neu
A:Reference number: A49069; MUID:94006554
A:Accession: A49069

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-772 <LUO>

A:Cross-References: GB:002528; NID:9410078; PIDN:AAC59638.1; PTD:9410079
C:Superfamily: semaphorin

Query Match 39.2%; Score 1862; DB 2; Length 772;
Best local similarity 46.8%; Pred. No. 4.6e-135;
Matches 370; Conservative 135; Mismatches 245; Indels 40; Gaps 12;

QY	103	WLG-----GILLHGGSSPSPGSPVPRRLRLSYRDLISNRSAIFLPGGSLNLAQMYL	156
DB	3	WLGILLSLGVLLAGRVNQHVNKNVPRKLISKLELSNINVTNGLANSSYHFFLL	62
QY	157	DEYDRFLFGGLDALYSLRLDQAMPREVLPPOGQRECVKGRPLTECANFERYL	216
DB	63	DEKRSRLVGAKHIFSENLVNI-KEYQKTVVSHSRDECKWAGKDILRECANFIKVL	121
QY	217	QPNRTHLLACGTAFQPTCALITVG-HRGEHYLLEPGSVESGRCRCPHEPSRPFASF	275
DB	122	KTVNQTHLYACGTAFHPMCTYLEVGSHPDNIHFEMDSHFENGKSPYDKLLASL	181
QY	276	IDGELTYGLTADFLGRAMIFRSGGPRPALRSDS-DQSLAHDPFVAAATLPENSDDND	334
DB	182	VDEELVSGTADFGRCFAFLTGHHHPIRTEDHDSRWMLNDPFIISAHLLIPESDNEED	241
QY	335	KYVFESSEIVSPDG--GSNVTVSRYGVNCVNDAGGQVLYNKKSTFLKARLYCVSGPG	393
DB	242	KYVFEEEN--AIDGHTGKATHTARIGICKNDGGRSLYNNKMTFLKARLYCVSGPN	299
QY	394	GAETHFDLEDVFLPWKAGKSLVYALFESTVASAVPQGFVAVCYHMAADIMEVFNGPFAHR	453
DB	300	GIDHNFELDVFELFMSKDKRNPITYVGTFTSSNIFGSAVCMYSDVRRVFLGPFAHR	359
QY	454	DGPOHMGPRYGGKVPKPRPGVCPSPKMTAOPGRPGSKKDPDEVLOPAPAHPLFWFVRP	513
DB	360	DGPYQWVPYVPGRVPTPRPGCPSPKTS---GPDSTKDLDPDEVITPARSHRANYNVFP	415
QY	514	RHGPRVLVTKTHLAQQLHQLIVVDRAVEDGTIVTLFGTSGSVLKVALAAGSAAPEPEV	573
DB	416	INSPRIMIKTVDVYQFTQIVVDRAVEDGQYDMVFGIDISTVLKVVSIPEKTHLEDEV	475

```

Query Match          38.6%  Score 1832:  DB 2:  Length 772:
Best Local Similarity 46.4%:  Pred. No. 9.3e-133:
Matches 365:  Conservative 135:  Mismatches 251:  Indels 36:  Gaps 11:

QY 101 ICWLLGILLHGGSSGSPSPSPRLRLSTYRDLJSANRSAIFLGPQSLNLOAMYLDEYR 160
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 7 IACLPMGVLLTARANYANGKNVNPRLKLTYSKEMLSNNVITFENGLANSSVHTFLILDER 66

QY 161 DRLFLFGIDALYLSLDDOAMPDRPRVILMPPOGAECEVKKRGRLPLTCANFVRLQPHN 220
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 67 SRLYGAADHIFSEFLVNI -KDFQKIVMPVSYTRDECKMAGKDLKCANFIKYLEAVN 125

QY 221 RTHLLACGTGAFOPTCALITYGHRGE -HVLLEPESVYESGRCGRCHEBSPFASTFIDE 279
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 126 QTHLYACGTGAFHPICTVILEVGHHEEDIIFRLQDSHFENGSGKSPYDKLLTASLIDGE 185

QY 280 LYTGLTADFLGREAMIFRSGGRPRALNSDS -DQSLHDPFVMAARIPENSDDQNDKYVF 338
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 186 LYSTADAEFMRDPAIFRTLLDHNPIRTEQDSFWMLNDRPISAHILIPESDNPEDDKYVF 245

QY 339 FFEFTVSPDGC--SNMYTVSRGVGVCVNDGAGGRVLYNKKSTFLKRLVYSGVPGCAE 396
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 246 FFRE---NAIGGEHSGKATHARICQICNDGGGRSLVKNKTTFLKARLLISTSVGPGMID 302

QY 397 THEPDQLEDFLLMRKAKSLSEVALVSTSAVFEQGFACVYHMADI EVEVNGPRAHNDGP 456
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 303 THEPDQLEDFVLMNSDKDRNPITYGVFTSSNIIFGSAVNCAMISMDVARYRVLFGPAHNDGP 362

```

[illegible]

OY	254	GSVSYSGRCRCPHEPSRPFASTEFIDGLYTLTDFLCREMIAPRSGGPPRALSD-SDOS	312
OY	159	GRIDGKGKSPYDRHNAASVIVGEELVSGVADIMGRDPTIFRSIGQRSLRTEPHDSR	218
OY	313	LHDPFVMAARIPENSDDNDKYFFFESETVIVSPDGGSNHVIVSKRGRCVANDAGQRY	372
Db	219	WLNPKRFVKFWMIPESENPDQDTYFFFRRTAEADALGRLSRSRGQICRDYVGQRS	278
OY	373	LVNKSFFLKLRLVCSVPGGAEETHFDQLEDFVLMPKAGKSLSEVALSTVSAVQGF	432
Db	279	LVNKTTFKLRLVCSVPVGG-DFHFDQLDVFLLSSRHRHRLPLXAVHSTSSSLRGS	337
OY	433	AVCVYHMADIWENGPRAHDFGPOHMGPRYGKVPFRPRGVCPSKKTAAQGRPFSGTKD	492
Db	338	AVCVYSMDYARAFLLGPFARHKEGPMHOMVSYGQVRPVRPRGMCPSKTFG---TFSTKD	393
OY	493	YRDELFQARAHPLMEFVVRPRHGRPYLVKTHLAQOLHOIVYDVEAEDEGTVDYDFLEIGTD	552
Db	394	FPDDVIOGRNHPPLMYSVLPTGSRPLFLQVGNVYTFQTQIADNVAAADGHVDFLIGTD	453
OY	553	SGSVLYKVTALDAGSSAEDEEVLLELOLFPKVPYDITEMEISVKRQMLVYSGRLCAQRL	612
Db	454	VGTVLKTVISYKSGRSPSAEGLLLELHVFEDSAVTSKQISSKRHOIYVARSAAVDIAL	513
OY	613	HOCETVYGAACAECCLARPYCAMDAGACTHVRPSLGRKRRPRDRIHGNPALDQLGSOE	672
Db	514	HRCAAHGKVTCECLARPYCAMDQVACTRRQPS-ARRFRRODVRRNDPSTLCSGSSR	572
OY	673	EEAVGLVATWVYGTETHNSTFLECLPKSPQAAVRLILOPDEGSPDOYKTDERYLHBERG	732
Db	573	P---ALLHEHVFEGVEGSSAFLECEPRSLQARVEMTFPRACVTAHTQVLAEBEKTERARG	628
OY	732	LIFRLSPFDAGTYCTTLEHGFSTVYRLAVIVASQNLNF-PPRPKEEPPARGGL	791
Db	629	LILRLRRDSSVYLCAVAEAGFTPLRLRLSHVISAQAERLARAEANAAAPG----	684
OY	792	ASTPRKAMYKDLILOI-----GFANLPVDEYCEKRWCRGTTEGSGC-FRSRSKGQAR	844
Db	685	-----PKIWWRDFLOLVFPGGGGSANSLRM-----CRPPALQSLPLAESRRKGRNRR	731
RESULT 5			
158169			
semaphorin III - mouse (fragment)			
C.Species: Mus musculus (house mouse)			
C.Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999			
C.Accession: 158169			
R.Messenger: E.K.; Leonardo, E.D.; Shatz, C.J.; Tessier-Lavigne, M.; Goodman, C.S., et al.			
A.Title: Semaphorin III can function as a selective chemoattractant to pattern sensory			
A.Reference number: 158169; MUID:95267432			
A.Accession: 158169			
A.Status: preliminary; translated from GB/EMBL/DBJ			
A.Molecule type: mRNA			
A.Residues: 1-666 <RES>			
A.Cross-references: GB:LA0484; NID:g703189; PIDN:AAA79334.1; PID:g703190			

[illegible]

```

Db 121 ISAHLPESDNEDOKVYFFEREN--AIDGESHGKATHARIQOICKNDGCRSLVNMKT 178
OY TELKARLVCSVPGGATGTHDQLEDVFLMLPKAKSLLEYALSTVSAVFOGFAVCYH 438
Db 179 TELKARLVCSVPGGATGTHDQLEDVFLMLPKAKSLLEYALSTVSAVFOGFAVCYH 238
OY MADIVEVNGEPHNRGPOHONGPVGKVPFPPGVCPSKMAQOGRPGSGKADPEVL 498
Db 239 MSDVRVLLGPRYAHRDGNVQWVYQGRVYPRPCTCSKTG---GFDSTKDLPDVYI 294
OY QFARHPLMFVPRRRGRPVLVKTHLAQOLHOIYVDRVEADGTYDVIETGDSGVLYK 558
Db 295 TFASSHPRAMPVPRPNNRPIMIKTDVNYOFTQIYVDRVDAEDGQYDVMFICTDGVLYK 354
OY VIALOAGSGAPPEEYVLEOVFKVPTITEMEISVKROMLYVSRVLAQOLHOCETY 618
Db 355 VVSPVKEVHMLEVLEEMTVFREPITISAMELSTKQOOLIGSTAGVAQDLPLHRCYI 414
OY 619 GTACAECCCLADPYCAMDGASTHYRPSLGRFRRODIRGNPALQCL-----GQ 669
Db 415 GKACAECCCLADPYCAMDGASTHYRPSLGRFRRODIRGNPALQCL-----GQ 669
OY 670 SOEEBAVGLVATVYVGEHNSTELECPKSPQAAVRLDOPDEGPDQYKTDERVLHT 729
Db 474 SLEER-----IYGVENSSTLECPKSPQAAVRLDOPDEGPDQYKTDERVLHT 525
OY 730 ERGLFRLSRFDAGTYCTLEHGFSTQVYVRLVVAQDLMLFPRPKPREPARG 789
Db 526 EQGLLSTLQKDDGNVYCHAVEHGFMTLVLEVIDTHELELHKDDGDSKTKE 585
OY 790 GLASTPP--KAMYDIIQILGIFANLPRVDEYCEVWCRGTEGSGCFRSRSGKQARGS 847
Db 586 MSSSMPTSPQKVMYDFDQILNHPMLNTWDEFCQVWKDKRQ----RQRGHSGSGSSNK 641
OY 848 WAGLEGKKKSRVHAENHRRPREV 872
Db 642 WKHMDSKSKGNRRTHFERAPRSV 666

```

```

RESULT 6
G02173
semaphorin III family homolog - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 24-Sep-1999
C:Accession: G02173
R:Naylor, S.
submitted to the EMBL data library, October 1995
A:Reference number: G09275
A:Accession: G02173
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-753 <NAY>
A:Cross-references: EMBL:U08276; NID:g1061350; PIDN:AB18276.1; PID:g1061351
C:Superfamily: semaphorin

```

```

Query Match 34.98; Score 1638; DB 2; Length 753;
Best Local Similarity 44.58; Pred. No. 2.2e-119;
Matches 345; Conservative 127; Mismatches 258; Indels 46; Gaps 18;

```

```

OY 104 ILGGLLHGG-SSGSPG-----PSVRLRLSYRDLSSANSALILGOGSLNLOAMY 155
Db 2 LVAGLLMSLTLGAMPSPFTQDHLPRVRLSPFKLKTGTAFHFNPLNTTDDYILL 61
OY 156 LDEYDRLEFLGIDLALYSRLDQAMPDPREVLPPOGREGCVRKGRDPLTECANFVR 215
Db 62 KDEDHDMYVGSKDYVLSLDLHDINNEPLIITHAASPORIECVLSKDVNGEGNFVRL 121
OY 216 LQHNHTHLLACGTGAFQPTCALITYGHRG-EHVLHPESVSGRGRCPHEPSRPAST 274
Db 122 IOPWNTHTLYVCTGAYNPCKTYVNRGRADQYIFYLEPERLESKKCKCPYDKPLDTASA 181
OY 275 FIDGELYTGLTADFLGRAMAFIRSGGRPALRSDSQS-LLHDPREYMAARIFENSQDN 333

```

```

Db 182 LINEELIYACVYIDFEGTDAIFRTLGKQYAMRTDYQNSWMLNDPSTIAELLPSDAEND- 240
OY 334 DKVYEFSESTVPSPDGGSNHTVSVHGRVYCVNDAGQVRLVNMKSTFLKARLVCSVPG 393
Db 241 DKLYEFFERR--SABAPQSPAYARIIGRICLUNDDGHCGLVKNMSTFLKARLVCSVPG 298
OY 334 CAETHTFDQLEDVFLMLPKAKSLLEYALSTVSAVFOGFAVCYIMADIVEVNGEPFAH 453
Db 299 GIETHFDQLEDVFOOTDQVNRVYAVFTSSVFRSGAVCVYSMADIRMFNCPFAHK 358
OY 454 DGPQHQMGVPGGKVPFPPGVCPSKMTAOPGRFSTQDYPREVLOFARHPLMFVPR 513
Db 359 EGPNOQWMPFSGKMPYPRPGTCPG-TTPRS--MSTADYPREVINTMRSHPLQAYVP 415
OY 514 RHGRVLYKTHLAQOLHOIYVDRVEADGTYDVIETGDSGVLYALQAGSAPREY 573
Db 416 LQRRPLVVRGTGAPVRLTTLAVDQVSDAGREYVFLGTDRGTQKIVLYPK-DQGEML 474
OY 574 VLEELQVFKVPTITEMEISVKROMLYVSRVLAQOLHOCETYGTACAECCCLADPYC 633
Db 475 MLEEVYFKDPAVKTWTISKRQOLYVSAVGYTHLHRCQAGACADCCCLADPYC 534
OY 634 AMDGASTHYRPSLGRFRRODIRGNPALQCLG--SOEEBAVGLVATVYVGEHNS 691
Db 535 AMDGACACRYTAS-SKRSRRODVHGNPIQCRGFNSNAKNV-----ESVQGVAGSA 589
OY 692 TFLKARLVCSVPGGATGTHDQLEDVFLMLPKAKSLLEYALSTVSAVFOGFAVCYH 750
Db 590 AFLKQPSRQVATVYKWLPRQDPGR-RREIRAFEDFLTEGLLRLALQLSDRGVLYCTA 648
OY 751 LEHGFSTQVYVRLVVAQDLMLFPRPKPREPARGGLASTPPKAMKDIQLIGF 809
Db 649 TENNFKHVYTVQLVLRDVAHVALEPPLSMAPRPGAG---PRPPYQELAQLLAQ 704
OY 810 ANLPRVDEYCEVWCR--GTEGSGCFRSRSGKQARGSWAGLEGKKKSRVH 862
Db 705 PEVGLIHQYCGYWRHVPSPREAPGARSPREPQD-----KTRRNRRH 748

```

```

RESULT 7
I48744
semaphorin A - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
C:Accession: I48744
R:Puschel, A.W.; Adams, R.H.; Betz, H.
Neuron 14, 941-948, 1995
A:Title: Murine semaphorin D/collapsin is a member of a diverse gene family and creat
A:Reference number: I48744; MUID:95267431
A:Accession: I48744
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-748 <RES>
A:Cross-references: EMBL:X85990; NID:g854323; PIDN:CAA59982.1; PID:g854324
C:Genetics:
A:Gene: sema
C:Superfamily: semaphorin

```

```

Query Match 34.98; Score 1653.5; DB 2; Length 748;
Best Local Similarity 46.18; Pred. No. 4.9e-119;
Matches 358; Conservative 106; Mismatches 250; Indels 63; Gaps 17;

```

```

OY 83 GRGRCPQPPSMASAMALQMLG-GILLHGSSGSPGSPVRLRLSYRDLSSANSAL 141
Db 2 GRAEAAVMDGL-----ALLVAGLG-----DTAENLPLRLSPQE-LQARHGRV 45
OY 142 FLPGOSLNLQAMVLDYDRLEFLGIDLALYSRLDQAMPDPREVLPPOGREGCVRK 201
Db 46 TFLKARLVCSVPGGATGTHDQLEDVFLMLPKAKSLLEYALSTVSAVFOGFAVCYH 105
OY 202 GRDPLTECANFVRLQPHNHTHLLACGTGAFQPTCAL---ITVGRHGEHVLHPESVSG 258

```


Db 106 GKDIGTECMNFVRLHVAHNTHTLLACRTGAFHPICALMRMTAG--GTGA-STGPKLEED 162
 QY 259 GGRCPHEPSPERASTFIDGLYGLTADFLGREAMTRSGPPALRS-D-SDGSLHDP 317
 Db 163 GKGTTPYDPRRPPSVLGEELYSVTDADLMGRDFTFRSLGONPSLTETPHDSRWLNEP 222
 QY 318 RYVMAARIPENSDDNDKVFYFFESFETVSPDGGSNHVTYSVRGVCVNDAGQAVLYNKM 377
 Db 223 KFYVFWIPESENDDDDTYFFEFRESAVEAARPMGRMSVSRVGOICRNDLGGQRLVYK 282
 QY 378 STFLKARLVCSVPFGAETHFDLEDFLLMPKAKSLEYALFSTVSAVFGFVAVCY 437
 Db 283 TTFLKARLVCSVPFGAETHFDLEDFLLMPKAKSLEYALFSTVSAVFGFVAVCY 437
 QY 438 HMADIWEVNGPFAHNRDQOMQGVGKVPFPPRGVCPGSKMTAOPRGFGSTKRYXPEV 497
 Db 342 SMNDVRAFLPRLPRHEKQPTQWVSQYGRVPRPGMCPKFTG---TFSTKDPDDV 397
 QY 498 LOFARHFLMPVPRPRGRPLVKTHLAQOLHQIVVDVREAEDEGTVDVIFLGTSGSVL 557
 Db 398 IOFGNHPFLMVPVLPMGGRPLFLOVAGVYFTQIADRVAAADHIDVLFIGTVGYVL 457
 QY 558 KYIALQAGSAPEEVEVLEELQVFKVPRPTTEMETSVKROMLYVSGRLGVAOLRLHOCET 617
 Db 458 KYISVPRGPRNSGILLLELOVFEDESAITSMOISSKROOLYVSAARAVALIALHRCIA 517
 QY 618 YGTAAECCIAADPCANDGASCTHYRPSLGRRRRRODIRGNALOCLOGSOEEAVG 677
 Db 518 IGRACAECCIAADPCANDGASCTHYRPSLGRRRRRODIRGNALOCLOGSOEEAVG 677
 QY 678 IYAAVMTSTENSTFELCPKSPQAAVFWMLQRPDEGPDQKTDERYLHTERGLLFRR 737
 Db 573 VLEKKVLGVEGSAFLCEPERSLQAHVQWTFQAGAEAHQVLAEEVERETARGLLHG 632
 QY 738 ISRFDACTYTTCTTLEHSGSQIVYRALVYASQDLNLFPRPKREPRPARGLASTP-- 795
 Db 633 LRROSGVYLCAVADGQFSQPLRLRLVHLVLSAQAERL-----ARAEAAAPAP 681
 QY 796 --PKMYRKDIOLI-----GFANLPVDEYCERWMCRTT--ECGCGFRSRSRQKAR 844
 Db 682 PPKMYRPLDQLVERGGGGGANSIRM-----CRPGHHSVADSRKRNRR 730

RESULT 8

semaphorin E - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
 C:Accession: I48748
 R:Puschel, A.W.; Adams, R.H.; Betz, H.
 Neuron 14, 941-948, 1995
 A:Title: Murine semaphorin D/collapsin is a member of a diverse gene family and creates
 A:Reference number: I48744; MUID:95267431
 A:Accession: I48748
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-751 <RES>
 A:Cross-references: EMBL:X85994; NID:g854331; PIDN:CAA59986.1; PID:g854332
 C:Genetics:
 A:Gene: semaphorin E
 C:Superfamily: semaphorin

Query Match 33.0%; Score 1568; DB 2; Length 751;
 Best Local Similarity 42.4%; Pred. No. 1.9e-112;
 Matches 337; Conservative 116; Mismatches 265; Indels 76; Gaps 15;

QY 100 AICWLGGLLH--GGSSGSPSPSPRLRLSTRDLSANRSLIFGPGSINLQWYVL 156
 Db 5 AICLVGVFICISICVRSQPO---ARVYLTFDELREKTSYFSLSHQOQDYLILM 59
 QY 157 DEYDRFLGLGLALYLRDQAMPDPREVILMPQPGQRECVKRGDPLTECANFVRL 216
 Db 60 DEODRIYVSKDHILSLNININISQEPVLSVFWPASTIKVECKNAGADPTGCGNFVRI 119

QY 217 QPHNRTHLLACGTGAFQPTCALITVGHGEHVLHLEPESVSGRCRCPHEPSPASTFI 276
 Db 120 QTFNRTHLYVCGSAGFSPVCYIANGRSEDEQVFMIDKCEGKRCGFENPVNTVSMI 179
 QY 277 DGEIYGLTADFLGREAMIFRSGRPPALRSDDS-LIHDRFVMAARIPENSQDNDK 335
 Db 180 NEELFSGMVIDFMGDDAIFRSLTRKMLQRLTDHNSKMLSEMFDAVHIDGTDPNAK 239
 QY 336 VYFFESFVPPSDGSSNHVTYSVRGVCVNDAGQAVLYNKMSTFLKARLVCSVPFGA 395
 Db 240 VYFFERKRLTDNNRSTKQIH-SMARICPDNTGGORSLYNKMSTFLKARLVCSVDECP 298
 QY 396 ETHFDLEDFVLLMPKAKSLEYALFSTVSAVFGFVAVCYHMADIWEVNGPFAHNRD 455
 Db 299 ETHFDLEDFVLLMPKAKSLEYALFSTVSAVFGFVAVCYHMADIWEVNGPFAHNRD 455
 QY 456 POHQMGPRGKVPFPPRGVCPGSKMTAOPRGFGSTKRYXPEV 497
 Db 359 PNHOLISTYOGRIPIPRPGTGPFGAFTPNMR---TTKDEPDVVTFLRNHPLMNSTISPLH 415
 QY 516 GRPLVKTHLAQOLHQIVVDVREAEDEGTVDVIFLGTSGSVLKYIALQAGSAPEEVL 575
 Db 416 RPLVIRIGTDYKTKIANDRVNAADGRHVLFLGTRGTQKVVVLPINSSAS-GEILL 474
 QY 576 EELQVFKVPTTEMETSVKROMLYVSGRLGVAOLRLHOCETYGTACACCLARDPYCAM 635
 Db 475 EELQVFKVPTTEMETSVKROMLYVSGRLGVAOLRLHOCETYGTACACCLARDPYCAM 635
 QY 636 DGASCTHYRPSLGRRRRRODIRGNALOCLOGSOEEAVGLVAADMYVTEHSTFLE 695
 Db 535 DGHSQVRYPT-GKRRSRQDVHGNPLTQCRGFLN--KAYRAAEIYQGVANNSTFLE 591
 QY 592 CAKSPQASIKMLQDKDKRKEG---KLNRIIATSQGLIRSVODSPOGLHYCIATE 647
 Db 696 CLPKSPQAAVFWMLQRPD---EGPDQVTDERYLHTERGLLFRRLSRDAGTYTTTLE 752
 QY 753 HGFSGTVRLALVYI-----VASQDLNLF-----PEPKREPRPARGLASTPPKMYK 802
 Db 648 NSFQKTIKINFEVLDSQEMVAVVTDKWSPWTWAGSVRALPFH-----K 652
 QY 803 ILQLIGFANLPVDEYCERWMCRTTECGCFRSRSGKQARCKSNAGLELGR-----K 856
 Db 693 ILGAFSHSEMQLINQY-----KTRQOQOOLGEPQKMRGDYK 731
 QY 857 MKSRVHAENRTPR 870
 Db 732 LKALINRSKRNR 745

RESULT 9

S66498
 M-sema F protein precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 24-Nov-1999
 C:Accession: S66498
 R:Inagaki, S.; Furuyama, T.; Iwahashi, Y.
 FEBS Lett. 370, 269-272, 1995
 A:Title: Identification of a member of mouse semaphorin family.
 A:Reference number: S66498; MUID:95385809
 A:Accession: S66498
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-834 <INA>
 A:Cross-references: EMBL:S79463; NID:g1110598; PIDN:AAB35184.1; PID:g1110599
 C:Superfamily: semaphorin
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-834/Product: M-sema F protein #status predicted <MAN>

Query Match 18.2%; Score 862; DB 2; Length 834;
 Best Local Similarity 31.6%; Pred. No. 4.0e-36;
 Matches 252; Conservative 106; Mismatches 282; Indels 158; Gaps 30;

QY 94 MASAIAICWLLGGLLHGSSGSPGSPVRLRLSLRDLISANSAIFLGPQSLNQA 153
 DB 1 MAFH-WAV-WLLAAGLWGLGIGAMMMNLVPRKTVSSSELTVVRRFSQTIQDFILITL 57
 QY 154 MYLDEYDRLFLGGLDLALYSRLDQ-----AMPDREVLVMPQPGORECVKRGDPL 206
 DB 58 --LTHSGLLVYGARELFAVSVALELQAIISWAPAE-----KKICTQKGSNQ 107
 QY 207 TECANFVRLVLPNHRTHLLACGTAFOPTCALITVGHGHEVHLLEPGESGRCPCHE 266
 DB 108 TECNFIREFLOPYNSSHLVYOCGYAFQPKCYIIM-----LTFLDRAEFEDGKCPDY 162
 QY 267 PSRPAFTFIDELGYTLADFLGREAMIFRSRGPRLRSDOSLLHDERVMAARIP 326
 DB 163 PARGHGTGLVDGLYSATLNNFLETCEPVLIRYMGTHHSIKTEYLAFWLNBEHFVGSFAVR 222
 QY 327 ENSDQ---DNDKYFEFSETPSPDGSNNHTVSRVGRVCVNDAGGQRLVKNKSTLKA 383
 DB 223 ESYGSFTGDDDKITFEFSENAVEIDCYSEOV-VARVARVCKGDMGARSATLQKWTTLKA 281
 QY 384 RLVCYVPGPGGAETHFDQLEDFVLLMPKAKSLLEVAFSTVASFOGFAVCVYHMADIW 443
 DB 282 RLVCASAP---DMKYEFNQLKAVHTLRGASWHNTTFEGVQARMGMDLSAVCEYQLEQIO 338
 QY 444 EVENGPRAHNDGQHOHMGPRYGKVPFRPGVCSKMTAQGRPGSTKDYDDEVYQARA 503
 DB 339 QVEGPRKEYSEQAOKKARYTDPVSPRPGSCJNNMHRDNG--YTSLELDPDNTLFIKK 396
 QY 504 HPLMFVPRRHRGRPVLYK-----THLAQOLHOIVDRAEADG--TYDVEFLGDSGV 557
 DB 397 HPLMEQVQRKRLGRPLLYKNTNFTH-----VADRVGLDGAITYTFLIGDQWLL 449
 QY 558 KYVALDAGGAPEEVEVLELOVKVPTPTIMEISVYKRLVYSGRLGVALRLHOCET 617
 DB 450 KAVSL-----GWIMHVELEQVFD--QEPVESIVLSQSKVLPAGSSQQLVQSLADCTK 502
 QY 618 YGACAECCCLARPYCAMDASCSTHRSYGRK-RPRRODIRHGNPALQCLGSGQSEAV 676
 DB 503 Y-REYDCVLAARPYCAMVNTSRCAVTSGRSGSEFLVQHVANLDTSKMC-----NOYGI 556
 QY 677 GLVAA-----TWYGTGHNSTFLECLPKSPQAAVRML-----ORPG--DEGPDQV 720
 DB 557 KYRSIRPKNITYVSGD---LVLPCHLSNLAHMTFSGDLPAPQSGFLDTG--- 609
 QY 721 KIDERYLHTEGGLFRRLSFDAGYTCTTLEHGSQTVVRLA---LVYIVA----- 769
 DB 610 -----LQALVMAAOSRHSRGPYRCYSEEG---TRIAESYLAVVAVAGSSVTLLE 655
 QY 770 --SOLDNL-----FPEE-PK-781
 DB 656 AARPLENLGLVWLAVALGAVCLVLLLVLSLRRLRELEKAKASERLYVPLELPKE 745
 QY 782 PEEPPARGGLASTPPKAW 799
 DB 716 PASPPFRPG-PETDEKLM 732

RESULT 10

148746

semaphorin C - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 03-Nov-2000

C:Accession: I48746

R:Puschel, A.W.; Adams, R.H.; Betz, H.

A:Title: Murine semaphorin D/collapsin is a member of a diverse gene family and creates

A:Reference number: I48744; MUID:95267431

A:Accession: I48746

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-782 <RES>

A:Cross-references: EMBL:X85992; NID:g854327; PIDN:CAA5984.1; PID:g854328

C:Genetics:

A:Gene: semc
C:Superfamily: semaphorin

Query Match

17.5%; Score 831.5; DB 2; Length 782;

Best local similarity 32.3%; Pred. No. 9.8e-56;

Matches 223; Conservative 110; Mismatches 281; Indels 77; Gaps 22;

QY 150 NLQAMYLDEYDRLFLGGLDLALYSRLDQAM---PDREVLVMPQPGORECVKRGDPL 206
 DB 15 NYTALLLSQDQKTLVYGARELFAVNSLFLPGGEYQELMSADADRKQCCSGKRPK 74
 QY 207 TECANFVRLVLPNHRTHLLACGTAFOPTCALITVGHGHEVHL-----EPGSV--ESG 259
 DB 75 RDCQNYIKILLPLNSSLHLLTGTAFSPLCYI-----HIAFTLAQDPAGVYILEDG 127
 QY 260 RGRCPHEPSRPAFTFIDELGYTLADFLGREAMIFRSRGPRLRSDOSLLHDERV 319
 DB 128 KGHCPFPDPRKSTALVVDGELTGTVSSFOGNDPAISRQSSRPT-KTESLNLQDPAF 186
 QY 320 VMAARFENSQ---DNDKYFEFSETPSPDGSNNHTVSRVGRVCVNDAGGQRLVKN 376
 DB 187 VASATSESLGSPIGDDDKITFEFSEGOEPEPENT-VSRVARVCKGDEGCBRLVQOR 245
 QY 377 WSTFLKARLYCSYVPGPGGAETHFDQLEDFVLLMPKA---GKSLVVALFTV--SAVEOG 431
 DB 246 WTSFLKAQLLCSRPDDG---FPFNVLDVFTLNNPDQMKRTLSI-GVFTSQHNRKTEG 301
 QY 432 FAVCVYHMADIWEYENGPRFAHRDGPQHOHMGPRYGKVPFRPGVCSKMTAQGRPGSTK 491
 DB 302 SAICVFETMNDVOKAFDGLYKVNRETOQWYETTHQVTPRPGACITNSARE--RKINSL 359
 QY 492 DYDEVLQFARHPLMPWRPRRHRGRPVLYKTHLAQOLHOIVDRAEADGTYDVEFLGT 551
 DB 360 QLPDRVLNFKDHFIMDGQVRSR-----LLLDPRARYQRAVHRVGLHSTYDVLGT 414
 QY 552 DSGSVLYKVALQAGGAPEEVEVLELOVKVPTPTIMEISVYKRLVYSGRLGVALRLH 611
 DB 415 GDRRLHRAVTLSS-----RVHIIIEELQIFPOGQPVQNLHDSHGLLVYSSHGAVQVP 468
 QY 612 LHOCETYGTACACGLLARDPYCAMDASC---THYRSLSKRRFRRODIRHGNPALQCLG 668
 DB 469 VANCSLYPT-CGCLLARDPYCAMTSGCARLASYLDYODLASRW-7QDIEGASVYKELCKN 526
 QY 669 QSGEEAVGLVAATMYG-----TEHSTFLECLPKSPQAAVRMLQRPDEGPDQV 720
 DB 527 SSYK-----ARFLVPGKPKQVOIQPNTVNTLACPLSLNATRLVY---HNGAPVNA 575
 QY 721 KIDERYLHTEGGLFRRLSFDAGYTCTTLEHGSQTVVRLALVYVASQDLNLEPPEP 780
 DB 576 SASCRVLPF-GDLLLVGSGQGLGVQCWSIEBGFQOLVASYCPVMEBEGVMDQKQMDG 633
 QY 781 KP-----EEPPARGGLASTPPKAWKDL 804
 DB 634 TVPIINTSRVSNAPAGRDSMGADKSTWNEFL 664

RESULT 11

148745

semaphorin B - mouse

C:Species: Mus musculus (house mouse)

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Nov-1999

C:Accession: I48745

R:Puschel, A.W.; Adams, R.H.; Betz, H.

A:Title: Murine semaphorin D/collapsin is a member of a diverse gene family and creat

A:Reference number: I48744; MUID:95267431

A:Accession: I48745

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-760 <RES>

A:Cross-references: EMBL:X85991; NID:g854325; PIDN:CAA5983.1; PID:g854326

C:Genetics:

A:Gene: semB

C:Superfamily: semaphorin

Query Match 12.8%; Score 606; DB 2; Length 724;

Best Local Similarity 29.0%; Pred. No. 2e-38; Matches 187; Conservative 94; Mismatches 260; Indels 104; Gaps 26;

QY 147 GSNLQAMVDEYDRFLGLDLYLRDQAMPD--PREVL-WPPOGORECVRKGR 203
 DB 64 GKLYYRFHNMEDDTLYGAMDRVFRVNLONISSNCRDAINLEPRDDVSCVSGK 123
 QY 204 DPLETCANFVRLDPHNR-THILACGTGAEPQ-----TCALITVHGRGHNHLEPGSVES 258
 DB 124 SQIFDCKNNHVRVIOSMOGDRLYVCGTNAHNPKYVYANLTHLPREXYV-----GVGL 178
 QY 259 GRCRCHEPRPRPASTFIDG-----ELYTGLADFLGREAMIFRSGRPALRSDS-- 309
 DB 179 GIACPCPDLDNSTAIYVNGNPGCLGSLGTAETKADTVIFRD-----LYNLSAK 233
 QY 310 -----DQSLIHDRFVMAARIPENSDODNDKVFYFSEF--VPSPDGGSNHYTV 356
 DB 234 RLETKKRTIKLYSKWLDKNEFSGSEFIDG-----YVFFEFRETAVEXYINCGR--AVY 284
 QY 357 SRVGRVCVNDAGGQRYLVNKMSTFLKARLYCVSPGCAETHFDQLEDVFLMPKACKSL 416
 DB 285 SRLARVCKKQVGGKNLLAHMMATYLRKALNCISGE--FFPYFNEIOSVYQL--PSDKS- 339
 QY 417 EYVALSTVSAVFOGPAVCVYHMAIWEVFNPGFAHRDGFQHOMGP-YGSKVPEPRGVC 475
 DB 340 REFAPFTTSTNGLIGSASFHINETIOAENKGFKEGSSNSAMLPVLSRVEPRGTC 399
 QY 476 PSKMTAOPGPFESTKDYDEVLQFARHPLMPVPRGRGRVLYKTHLAQOLHOIIVD 535
 DB 400 -----VNDISNLPDYVNTFTRSHPLMDKAVNHEINNIVYKRDLY--FTKLIWD 446
 QY 536 ---RVEAEDGTVDVLEIGTDSGLKAYIALQAGSAPDEEVELELOVFKPR--TPITENE 691
 DB 447 KIRIDILNQEYIYYVYGTNIGRIYKIVQYRNG-----ESLSKLIDIFEVAPREALQVNE 501
 QY 592 ISVKRQMLVGSRLGAQOLRIHOCETGYTACAECCCLARDPYCAMD--GASCHYRPSLGR 649
 DB 502 ISQTRKSLTIGDHRKIQIDILAMCNRRYDNCFRG--VRDPYCGMKDEANTCRPEYEDL-- 557
 QY 650 RFRFRDRIHGNPALQCLQSOEEEAVALAATMVYCTEHNSTFLECLPKSPQA-----AV 705
 DB 558 -----LQDVANETSDICDSSVLLKKIYVITGOSVHLGCEVYKPIEVLKNEGV 603
 QY 706 RWLLQRPQDEGPDQYK--TDERVLTHERGLFRRLSRPDAGTYTC 748
 DB 604 TW-YHNSKDKRGYELIRYPTKYTEETTERGLVYVSVNEADGGRYDC 647

RESULT 14
 B49423
 semaphorin I - fruit fly (*Drosophila melanogaster*) (fragment)
 C:Species: *Drosophila melanogaster*
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 07-May-1999
 C:Accession: B49423
 R:Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.
 A:Title: The Semaphorin genes encode a family of transmembrane and secreted growth cone
 A:Reference number: A49423; MUID:94094332
 A:Accession: B49423
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-656 <KOL>
 A:Cross-references: GB:I26082
 C:Genetics:
 A:Gene: semat
 A:Cross-references: FlyBase:Fbgn0011259

Query Match 12.7%; Score 601.5; DB 2; Length 656;
 Best Local Similarity 29.2%; Pred. No. 3.9e-38;
 Matches 190; Conservative 94; Mismatches 272; Indels 95; Gaps 24;

QY 208 ECANFVRLQPHNRTTHLLACGTGAFOPTCALITVGRGHNHLEPGSVESGRCHPEP 267
 DB 3 DCONVIRIMVSPGRLEFVCGTNSFRPMCNTHIISD-SNTYLE-----ATKQCAVCPDP 57
 QY 268 SRPASTFIDGELYTGLADFLGREAMIFRSGRPALRSDSLSLHDPFRVMAARIPE 327
 DB 58 RHNSISVLADNELYSGLVADSGSDPIIYR--EPLOTQEQYDSLISNAFNV----- 106
 QY 328 NSDQDNKRVYFSEF--VPSPDGGSNHYTVSRVGRVCVNDAGGQRYLVNKMSTFLKARLY 386
 DB 107 SSFTQGDVEYFREFRETAVEFINCK--AIYSRAVARVCMDKGGHRRRNKRTSLKSLN 164
 QY 387 CSVPGCAETHFPOLEDFVFLMPKACKSLF--VVALFSTVSAVFOGPAVCVYHMAI 443
 DB 165 CSI--PEDYPRFENEIOSASNLVEGQYSMSKLIYGVFNTPNSISFGSAVCARALQDIA 222
 QY 444 EVENGPRFHRDGFQHOMGPY-GGKVPFPRGVCPSKMTAOPGRFGSTKDYDEVLQFAR 502
 DB 223 DTFEGQFKEDGTGINSNMLPVNNAKVPDPREGSC-----HNSRALPDFTLNFIR 271
 QY 503 AHPLEMPVRPRHGRVLYKTHLAQOLHOIIVD-RVEAEDG-TYDYLEIGTDSGLVLY 560
 DB 272 THSLMDENVAFESQPIVTRSTIYRFTQIAYDAQITPGKTYDVFVGTDRKTIKSV 331
 QY 561 ALQAGSAEP-EEVYLELOVFKVPYITLMEISVKRQMLYVGSRLG----- 606
 DB 332 NASADSDADKTVSVLEIDVLTKESEIRNLEI--VFTMQYDQPKDGSYDDGKLIITVS 389
 QY 607 -VAQLRIHOCETFG-YACAECCCLARDPYCAMD--GASCHYRPSLGRFRFRDRIHGN 661
 DB 390 QVVALQIHRCHNKKITSCSCVALADPYCAMDKIAGKNSHGAPRMLENEYQVATQ 449
 QY 662 PALQCLQSOEEEAVALAATMVYCTEHNSTFLECLPKSPQAARWLLQRPQDEGPD-- 718
 DB 450 HAACPSGKINSKDA---NAGECKGFRNDMLDSRQSKDQETIDINIDK--NEGQQTSA 504
 QY 719 -----QKTDERYLTERGLFRRLSRPDAGTYTCITLLEGEFQYVRLALVYVASQLD 773
 DB 505 DIINAQYVETILVMAVLASIFSLGVEFTG-YGCRRCG-----KDEDD 548
 QY 774 NLPPEPRKEEPPARGGLASTPRKAWYKDILQIGFANLPRVDE--YCERV 822
 DB 549 NLPYDPEYEFQFQONVNSPSSCRIOQEPKL-----LPQVEEYIYAPV 594

RESULT 15
 T27165
 hypothetical protein Y54E5B.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27165
 R:Lennard, N.
 A:Title: submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20321
 A:Accession: T27165
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-712 <WIL>
 A:Cross-references: EMBL:AL032653; PIDN:CAA21714.1; GSPDB:GN00019; CESP:Y54E5B.1
 A:Experimental source: clone Y54E5B
 A:Genetics:
 A:Gene: CESP:Y54E5B.1
 A:Map position: 1
 A:Introns: 36/1; 70/2; 85/3; 194/2; 221/1; 253/3; 301/1; 352/3; 522/2; 560/1; 599/3;

Query Match 12.6%; Score 599; DB 2; Length 712;
 Best Local Similarity 29.5%; Pred. No. 6.9e-38;
 Matches 196; Conservative 98; Mismatches 254; Indels 116; Gaps 31;
 QY 161 DRLFGLDLYSLRLDQAMPDPREVLPPOGORECVRKGRDPLTCANFVRLQPHN 220
 DB 63 DSLVAGANNAVYNLSLSTLSVN-HKIDMKPAEHIIEICIMGKSK-TDCQNYIRVLARKS 120

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: October 9, 2002, 13:14:58 ; Search time 24 Seconds
(without alignments)
1411.650 Million cell updates/sec

Title: US-09-813-290-2

Sequence: 1 MACALAGKVPKMGSPVWVK.....KKSRVHAHNRPREVAT 875

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2043.5	43.1	775	1	SM3E_MOUSE
2	2032.5	42.8	785	1	SM3E_CHICK
3	2014.5	42.4	775	1	SM3B_HUMAN
4	1862	39.2	772	1	SM3A_CHICK
5	1846	38.9	772	1	SM3A_RAT
6	1843	38.8	772	1	SM3A_MOUSE
7	1840	38.8	778	1	SM3B_MOUSE
8	1830	38.6	771	1	SM3A_HUMAN
9	1790.5	37.7	761	1	SM3D_CHICK
10	1770	37.3	860	1	SM3C_MOUSE
11	1766.5	37.2	764	1	SM22_BRARE
12	1747	36.8	749	1	SM3B_HUMAN
13	1721	36.3	777	1	SM3D_HUMAN
14	1653.5	34.8	778	1	SM3B_MOUSE
15	1652	34.8	785	1	SM3F_HUMAN
16	1643	34.6	785	1	SM3F_MOUSE
17	1568	33.0	751	1	SM3C_MOUSE
18	1561.5	32.9	751	1	SM3C_CHICK
19	1538	32.4	751	1	SM3C_HUMAN
20	884.5	18.6	837	1	SM4G_MOUSE
21	874.5	18.4	861	1	SM4D_MOUSE
22	867.5	18.3	834	1	SM4C_HUMAN
23	862	18.2	834	1	SM4C_MOUSE
24	861	18.1	838	1	SM4B_HUMAN
25	831.5	17.5	782	1	SM4B_MOUSE
26	825	17.4	766	1	SM27_BRARE
27	776	16.4	770	1	SM4F_HUMAN
28	759	16.0	777	1	SM4F_MOUSE
29	750.5	15.8	776	1	SM4F_RAT
30	706.5	14.9	776	1	SM4B_HUMAN
31	698.5	14.7	888	1	SM6B_HUMAN
32	693	14.6	886	1	SM6B_MOUSE
33	682	14.4	888	1	SM6A_MOUSE

34	678.5	14.3	887	1	SM6B_RAT	070141	rattus norv
35	676.5	14.3	760	1	SM4A_MOUSE	062178	mus musculus
36	664	14.0	931	1	SM6C_MOUSE	09wt13	mus musculus
37	646	13.6	930	1	SM6C_HUMAN	09wt13	homo sapien
38	645	13.6	960	1	SM6C_RAT	09wt13	rattus norv
39	628	13.2	730	1	SM1A_SCHAM	026473	schistocerc
40	627.5	13.2	771	1	SM1A_DROME	024322	drosophila
41	620.5	13.1	697	1	SM2A_SCHGR	09x268	schistocerc
42	599	12.6	712	1	SM1A_GABEL	017330	caenorhabdi
43	588	12.4	706	1	SM2A_DROME	024323	drosophila
44	571.5	12.0	1093	1	SM5B_MOUSE	060519	mus musculus
45	555.5	11.7	1077	1	SM5A_MOUSE	062217	mus musculus

ALIGNMENTS

RESULT 1
ID SM3E_MOUSE STANDARD: PRT; 775 AA.
AC P70275; 009078; 009079;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 3E precursor (Semaphorin H) (Sema H).
GN SEMA3E OR SEMAH OR SEMH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=98175564; PubMed=9515811;
RA Christensen C.R.L., Klingelhoefer J., Farabykina S., Hultgaard E.F.,
Kramerov D., Lukanidin E.
RT "Transcription of a novel mouse semaphorin gene, M-semah, correlates
with the metastatic ability of mouse tumor cell lines";
RL Cancer Res. 58:1238-1244(1998).
RN [2]
RP REVISIONS.
RA Christensen C.R.L.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN EMBRYONIC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: DEVELOPING LUNGS, DEVELOPING SKELETAL
ELEMENTS, AND VENTRAL HORNS OF THE DEVELOPING NEURAL TUBE.
CC -!- CORRELATES POSITIVELY WITH TUMOR PROGRESSION.
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
EMBL: Z80941; CAB02590.1; -
EMBL: Z93947; CAB07987.1; ALT-SRD.
EMBL: Z93948; CAB07988.1; ALT-SRD.
DR EMBL: Z93948; CAB07988.1; ALT-SRD.
DR MGI:1340034; Sema3e.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PR00047; Ig_1.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00409; Ig; 1.
DR SMART: SM00423; PSI; 1.
KW Signal; Immunoglobulin domain; Multigene family; Neurogenesis;

Query Match 42.8%, Score 2032.5; DB 1; Length 785;
 Best Local Similarity 50.6%; Pred. No. 9.2e-143;
 Matches 398; Conservative 120; Mismatches 236; Indels 33; Gaps 12;

100 AICWLLGGLLHGSSGSPSPSVRLRLSYRDLISANRSAIFLPGSLNQAAYLDEY 159
 15 ALC---GLLELPAGYHATDTRQPLRLSHKELMDLNTSVFHSFGLHIMLDEY 70
 160 RDLRLGGLDALYSRLDQAMPDPPEVLMPPQPGORECEVGRDPLTECANFVRLQPF 219
 71 QERLEFVGGRDLISLSDRISNNYHEIHPSTPLAEEIIGRD-ADBCANVYVRLHXY 129
 220 NHTHLLACGTGAFDPTCALITYGHRG-HVHLPEGSVESGRCRCPHEPSRPASTFIDG 278
 130 NHTHLLACGTGAFDPTCALITYGHRG-HVHLPEGSVESGRCRCPHEPSRPASTFIDG 189
 279 ELYTGLTDFGREGMIFRSRSGRPALRSDD-OSLLHDPFRVMAARIPENDODNDKY 334
 190 ELFTGLSDYMGKDAVAFPTKRMMAHLKTEPDSEHLKPKFVGSYMPIDNEDHDNDKY 249
 338 FFESETVSPDGSNNHVVYSRYGVCVNDAGQRYLVNKMSTFLKARLYCSVPGGAGT 397
 250 LEFTFKALEAF-TSTHAIYTRYGRVCVNDMGQRIYVNMKSTFLKARLYCSVPGRGIDT 308
 398 HEDQLEDFVLLMPKAKSLVYALFSTVSAVFOGAVCVYHMAIDWEVNGPPAHRDGPQ 457
 309 HDELEDVYLLQTRDNKNVIFGLFSTSNIFRGYALICYHMAIYVAAENGPAHKEGPE 368
 458 HMGVYGGVPPRRPGVCSKMTAOPRGFGSTKDYPRDELOFARHPLMFVPRRHR 517
 369 YWALYEGVPPRRPGVCSKMTAOPRGFGSTKDYPRDELOFARHPLMFVPRRHR 425
 518 PVLVTHLAQQLHQAIVVDVEAEDGTVDYIFLGTDSGLYKIALQAGSAPPEVLEE 577
 426 PLVATDQKYNLKOIYADVEAEDGQYDLFIGTONGYLVKVTITINQETESMEVILDE 485
 578 LOVFVPPVPTITMEISVSKOMLYVSRGLVQAOLRHQCTGTACAECLARDPYCAMG 637
 486 LOVFVPPVPTITMEISVSKOMLYVSRGLVQAOLRHQCTGTACAECLARDPYCAMG 545
 638 ASCTHYRPS--LGRFRERODIRHGNPALOCGOSQEEVAVLVATVYGTNHNSTPLE 695
 546 ICSRYPPGMAKRRFRKQVHGNAAQCCFGQOFTGEVLEKTERLYVGLTYNSTLLE 605
 696 CLPSPOAAVRLMLODEGEPDOVKTDERVLTERTGLFRRLSRPDAGTYCTTCTENHG 755
 606 YPRPLQAAVNNFVQAHHTKKEEVYTERIITKMDGLFLFKHRLADAGYFCQYVHST 665
 756 SOTVVALVYVASOLDNLPPEPKPE---EPPARGGLASTPPKAMYKDLILIGRA 810
 666 VHTVRKITEIYEEERVDMSKDYEEIISHKMPCCMOGINPOV-SKPMYKEFLILIGYS 724
 811 NLPVDEYGERWACRTGCSGCFRSRSGKOARGKSMAGLGGK----MSRVHAH 865
 725 NQORVEIEEKWCID-----KRRKKLKMSPK-WKYANPOEKRODEKAKIRPEH 775
 866 NRTPREV 872
 776 YRLPRNI 782

RESULT 3
 SM3E_HUMAN STANDARD; PRT; 775 AA.
 AC 015041;
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Semaphorin 3E precursor.
 GN SEMA3E OR KIA00331.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,
 RT Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.
 RL DNA Res. 4:141-150(1997).
 RN [2]
 RP SEQUENCE OF 1-38 FROM N.A.
 RA Holmes A., Johnson D., Elliot G., Burkhardt J.,
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC [3]
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: AB002329; BAA20789.1; -
 DR EMBL: AC004954; AAC69513.1; -
 DR InterPro: IPR003589; Ig.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SM00409; Ig; 1.
 DR SMART: SM00423; PSI; 1.
 DR Signal: Imm0010bulin domain; Multigene family; Neurogenesis;
 KW Developmental protein; Glycoprotein.
 FT SIGNAL 1 25
 FT CHAIN 26 775
 FT DOMAIN 241 540
 FT DOMAIN 647 736
 FT DOMAIN 737 770
 FT DISULFID 654 729
 FT CARBOHYD 44 44
 FT CARBOHYD 126 126
 FT CARBOHYD 330 330
 FT CARBOHYD 595 595
 FT CARBOHYD 596 596
 SQ SEQUENCE 775 AA; 89227 MW; CD6079C1AE48F779 CRC64;
 Query Match 42.4%, Score 2014.5; DB 1; Length 775;
 Best Local Similarity 50.3%; Pred. No. 1.9e-141;
 Matches 395; Conservative 130; Mismatches 236; Indels 25; Gaps 10;

94 MARSANALICWLLGGLLHGSSGSPSPSVRLRLSYRDLISANRSAIFLPGSLNQA 153
 1 MASAGHITLTLGWLGLLELTGCGHTADTHPRRLSHKELINLRNTSIFSPGFLDLHT 60
 154 MYDEYRDLFLGGLDALYSRLDQAMPDPPEVLMPPQPGORECEVGRDPLTECANFV 213
 61 MLDEYQERLFLVGGRLVYSLSLERISDGYKEIHPSTALKMECIKKGD-AGECANYV 119
 214 RVLQPHNRTHLLACGTGAFDPTCALITYG-HRGHVLHPEGSVESGRCRCPHEPSRPA 272
 120 RVLHNYRTHLLACGTGAFDPTCALITYG-HRGHVLHPEGSVESGRCRCPHEPSRPA 179
 273 STEFDGLYGLTADFLGREGMIFRSRSGRPALRSDD-SIDSLHDPFRVMAARIPENDQ 331
 180 STILGSLFAGLVSDYMGKDAVAFRSMGRIAHITETEDDRLLKPKFVGSYMPIDNEDR 239

QY 332 DNDKVFEEFSETPDGGSNHVTVSRYGVGVNDAGORVLVNMKSTFLKRLVCSVPG 391
 CC - SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DB 240 DNDKVFEEFSETPDGGSNHVTVSRYGVGVNDAGORVLVNMKSTFLKRLVCSVPG 398
 QY 392 PGAEETHEDQEDVFLMPKAGKSELYVALSTVSAPVQGAFCVYHMADIWEVFNPPA 451
 CC - SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DB 299 MNGIDTFDELEDEVLPTDRHKNFVIEGLFNTSTNIFGHAICVYHMSISRAAFNGPYA 358
 QY 452 HRDQFQHWGKGVKVPVPRGVCSKMTAQRGPRGSKDVPDEVLQARAHPLMFVWP 511
 CC - SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DB 359 HKEPEPEHWSYEGKVPVPRGVCSKMTAQRGPRGSKDVPDEVLQARAHPLMFVWP 511
 QY 512 RPRGRPLVYTHLAQOLHQTIVDVEAEADGTVDYIFLTGDSGVLYKVALDAGSABEP 571
 DB 416 KPAHKPLVLTGDKYVNLKQIADVEAEADGTVDYIFLTGDSGVLYKVALDAGSABEP 475
 QY 572 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 DB 476 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 QY 632 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 DB 536 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 QY 690 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 DB 596 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 QY 750 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 DB 656 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 QY 805 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 DB 715 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 QY 865 HNRTPR 870
 DB 765 HNRTPR 870

RESULT 4
 SMA_CHECK STANDARD; PRT; 772 AA.
 AC 090607;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Semaphorin 3A precursor (Collapsin-1) (COLL-1).
 GN SEMA3A OR COLL1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 322-329; 362-372; 395-403 AND
 RP 666-680.
 RC TISSUE=Brain;
 RX MEDLINE=94006554; PubMed=8402908;
 RA Luo Y., Raible D., Raper J.A.;
 RT "Collapsin: a protein in brain that induces the collapse and paralysis
 RT of neuronal growth cones";
 RL Cell 75:217-227(1993).
 CC - FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
 CC CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
 CC BY A MOTILITY-INHIBITING MECHANISM. BINDS TO NEUROPILIN.
 CC - SUBCELLULAR LOCATION: Secreted (by similarity).
 CC - TISSUE SPECIFICITY: EXPRESSED AT RELATIVELY HIGH LEVELS IN BRAIN
 CC AND MUSCLE, MODERATE LEVELS IN LUNG, BURSAS, AND HEART AND
 CC VIRTUALLY ABSENT IN LIVER. COLLAPLIN-1, -2, -3, AND -5 BIND TO
 CC OVERLAPPING BUT DISTINCT AXON TRACTS.
 CC - DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY

CC THIRD OF THE PROTEIN.
 CC - SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DB 512 RPRGRPLVYTHLAQOLHQTIVDVEAEADGTVDYIFLTGDSGVLYKVALDAGSABEP 571
 QY 572 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 DB 476 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 QY 632 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 DB 536 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 QY 690 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 DB 596 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 QY 750 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 DB 656 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 QY 805 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 DB 715 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 QY 865 HNRTPR 870
 DB 765 HNRTPR 870

Query Match 39.2%; Score 1862; DB 1; Length 772;
 Best Local Similarity 46.8%; Pct. No. 3.8e-130;
 Matches 370; Conservative 135; Mismatches 245; Indels 40; Gaps 12;

QY 103 WLIG-----GLLHGSSPSPSPSPRLRLSLRDLISANRSAIFLPGQSLNQLQAYL 156
 DB 3 WLIGLALLSGVLLAGRVNCOHQVKNPRLKLSYKEMLESNNIVNFGNLANSSTYHTFL 62
 QY 157 DEYRDLFLGGLDLYSLRDLQAMPDREVLMPQPGQRECVAKKGDPLTECNFVRL 216
 DB 63 DEYRDLFLGGLDLYSLRDLQAMPDREVLMPQPGQRECVAKKGDPLTECNFVRL 216
 QY 217 QPHNRTHLLACGTAFOPTCALITYG--HREGHVLTLPESVSGRGCPHEPRPFASTE 275
 DB 122 KTYNQTHLYACGAGFHPMCTYIEVSGHPEDNIFRMEDSHFENGGRKSPYDKLLTASLL 181
 QY 276 IGCGLTGLTADLGLGEMAFRRSGGPRALRSOS--DOSLHDPFVMAARIPENSODND 334
 DB 182 VQGLYSSTGAADPMGDFALFRLTGHHPHIRTBOHOSRWLNDPRFISAHLPESDNPEDD 241
 QY 335 KYVFESEVTPSPDG--GSNHTVTSRYGVGVNDAGORVLVNMKSTFLKRLVCSVPG 393
 DB 242 KYVFESEVTPSPDG--GSNHTVTSRYGVGVNDAGORVLVNMKSTFLKRLVCSVPG 393
 QY 394 GAETHEDQEDVFLMPKAGKSELYVALSTVSAPVQGAFCVYHMADIWEVFNPPA 451
 DB 299 MNGIDTFDELEDEVLPTDRHKNFVIEGLFNTSTNIFGHAICVYHMSISRAAFNGPYA 358
 QY 452 HRDQFQHWGKGVKVPVPRGVCSKMTAQRGPRGSKDVPDEVLQARAHPLMFVWP 511
 DB 359 HKEPEPEHWSYEGKVPVPRGVCSKMTAQRGPRGSKDVPDEVLQARAHPLMFVWP 511
 QY 512 RPRGRPLVYTHLAQOLHQTIVDVEAEADGTVDYIFLTGDSGVLYKVALDAGSABEP 571
 DB 416 KPAHKPLVLTGDKYVNLKQIADVEAEADGTVDYIFLTGDSGVLYKVALDAGSABEP 475
 QY 572 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 DB 476 EYVLEELQVFPVPTIMEISVKRQMLYVSRGLVQALRLHQCETYGTAECCLARDP 631
 QY 632 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 DB 536 YCANDGASCTHYRP--SLGKRFRFRDRIHGNPVALQCGOQSEEAAGVLAATMYGTE 689
 QY 690 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 DB 596 NSTFLKLPKSPQAVRWLQRPDDEGPDQVKTDERVLHTERGLLFRRLSRPDAGTYTCT 749
 QY 750 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 DB 656 TLEHGEQTVRLVLYVAVSOLNLPFPEKPEEP-----PARGGLASTPPKANKYKDL 804
 QY 805 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 DB 715 QLTGFANLPRVDEYCEWYRCGTETECSCFRSRSRGKQARGKSMAGLGLGKMKSRVHAE 864
 QY 865 HNRTPR 870
 DB 765 HNRTPR 870

```

DB 476 LLEEMTVREPTVISAMKISTKOOOLYIGSATVSOQLPLHRCDPVYGACAECLARDPYC 535
QY 634 AMGASCTHRPESLGRRRRRODIRHGNPALOC-----LGOSOEBAVGLVATWY 684
DB 536 AMWSSCSSTFPT-AKRRTRRDRIKNGDPLTHCSDLQHDHNPSCQTLLEER-----IT 586
QY 685 YGTEHSTFTECLPKSPQAVRWMLORPDEGPDQVYTERVLTENGELFRRLSRDAG 744
DB 587 YGTEHSTFTECLPKSPQAVRWMLORPDEGPDQVYTERVLTENGELFRRLSRDAG 646
QY 745 TYTCTLEHGFSTQVYRLVIVASQDLNLFPPPEKPEEPARGGLASTPP--KANYKD 802
DB 647 IYFCHAVEHGFIOITLKVLTLEVIDTDLHELLHKEEDADASKTKDATNSMTPSQKIWYRD 706
QY 803 ILDILGFANLPREDCECERWCGTTECCSCFPRSNGKQARGSMAGLELKKMKSRVH 862
DB 707 EMOLINHPNLTNDEFCEOVWKDKRKO----RROPANAQVNTNKKHLDENKKGRRRT 762
QY 863 AEHNRTPREV 872
DB 763 HEFERAPRSV 772

RESULT 5
SM3A_RAT STANDARD: PRT: 772 AA.
ID 063548:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 3A precursor (Semaphorin III) (Sema III).
GN SEMA3A.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=MISTAR; TISSUE=Brain;
RX MEDLINE=97073089; PubMed=8915837;
RA Giger R.J., Wolfer D.P., De Wit G.M.J., Verhaegen J.;
RT "Anatomy of rat semaphorin III/collapsin-1 mRNA expression and
RT relationship to developing nerve tracts during neuroembryogenesis.";
RL J. Comp. Neurol. 375:378-392(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN GUIDING GROWING AXONS TOWARDS THEIR
CC TARGETS BY FORMING A MOLECULAR BOUNDARY THAT INSTRUCTS AXONS TO
CC ENGAGE IN THE FORMATION OF SPECIFIC NERVE TRACTS. BINDS TO
CC NEUROPILIN.
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- DEVELOPMENTAL STAGE: AT E-11, EXPRESSION WAS RESTRICTED TO THE
CC OLFACTORY PIT, THE BASAL AND ROSTRAL SURFACE OF THE TELECEPHALIC
CC VESICLE, THE EYE ANLAGE, THE EPITHELIUM OF RAHKE'S POUCH, AND
CC SOMITES. AT LATER DEVELOPMENTAL STAGES, IT WAS WIDELY DISTRIBUTED
CC IN NEURONAL AS WELL AS IN MESENCHYMAL AND EPITHELIAL STRUCTURES
CC OUTSIDE THE NERVOUS SYSTEM. AFTER BIRTH, MESENCHYMAL LEVELS
CC DECREASED RAPIDLY AND EXPRESSION BECAME RESTRICTED TO SPECIFIC
CC SETS OF NEURONS IN THE CNS. IN THE MATURE CNS, IT IS DETECTABLE IN
CC MITRAL CELLS, NEURONS OF THE ACCESSORY BULB AND CEREBRAL CORTEX,
CC CEREBELLAR PURKINJE CELLS, AS WELL AS A SUBSET OF CRANIAL AND
CC SPINAL MOTONEURONS.
CC -1- DOMAIN: STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: X95286; CAA64607.1;
DR InterPro: IPR003599; I9.
DR InterPro: IPR003006; I9_MHC.
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF00407; I9; 1.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00409; I9; 1.
DR SMART: SM00423; PSI; 1.
KW Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1..20
FT CHAIN 21..772
FT DOMAIN 240..538
FT DOMAIN 643..730
FT DOMAIN 728..770
FT DISULFID 650..723
FT CARBOHYD 53..53
FT CARBOHYD 125..125
FT CARBOHYD 591..591
SQ SEQUENCE 772 AA; 88808 MW; 240907812FF9F2D2 CRC64;

Query Match 38.9%; Score 1846; DB 1; Length 772;
Best Local Similarity 46.6%; Pred. No. 5.8e-129;
Matches 366; Conservative 138; Mismatches 248; Indels 34; Gaps 11;

QY 101 ICWLLGLGLHGGSSGPGSPVRLRLRYRDLISANRSLFGLPGSLNQLQAMLYDEYR 160
DB 7 IACLFWGLLFLARANYANGKNNVRLKISTYKMLESSNVITFNGLIANSSTHTFLDLDER 66
QY 161 DRLEFLGLDALYSRLDQAWDPREVIMPPQPGORECECVKRGDRPLTECANFVRYLOPHN 220
DB 67 SRLVYGADHIFSENLVNI-KDFOKIYVPVSYTRRDECKWMGKILKCANFIVLKAYN 125
QY 221 RTHLLAGTGAFOPTCALITTYGHRGE-HVLLHLEPGSVSSGGRCPHEBSRFASFIDGE 279
DB 126 QTHLYACGTGAFHPTCTYIEVGHNEEDNIFKLDOSHFEHNGSKSPYDKLLTASLDIGE 185
QY 280 LYTGLTADFLEAMIFRSGGPRPALRSDS-DQSLHPPRVMAARIPENSDDNDKYF 338
DB 186 LYSGTADPFMRGDRALFTTIGHNHDIRTEQHSRLNDRPRLISALIPESDNPEDDKYF 245
QY 339 FFSETVSPDG-GSNHVTVSRYGRVCVADGQRYLVNKMSTFLKARLVCSYPGGCAET 397
DB 246 FFREN-AIDESHGSKATHARIGQICKNDFGHRSVLWKTFTFLKARLVCSYPGGCAET 303
QY 398 HFDQLEDFLLMPKRAKSLLEYALFSTYSAYFOGAVGVYHMDIWEVFNCGFAHRDGPQ 457
DB 304 HFDQLEDFLLMNSKDPKMPYVGVFTTSNIFKGSAYCVMSDVRVFLGYAHRDGPN 363
QY 458 HOMGVYGGKVPFPPRGVCPKTAOPGRFGSTKDPDEVLOFARAHPLMFWYPRRGR 517
DB 364 YQWVYQGRVYPRPGTGPSTEG---GFDSTKLPDQVITFAFSHPAMNPVPIINR 419
QY 518 PVLVYTHLAQDLQIVDVREAEQTYVIFLGTDSGVLYKALVLAQAGSAPEEYVLEE 577
DB 420 PIMTIDVNYQFTQIVRVYDAEDQYDMFETGVTVLYKVVSPKKTMDLEELVEE 479
QY 578 IQVAVPPTITEMETSVKRWMLYVGSPLGVNQLRHQETGTACAECLARDPYCAMDG 637
DB 480 MIVFREPTTISAMELSTKQOOLYIGTAGVADPLHRCDIYKACAECLARDPYCANDG 539
QY 638 ASCTHYRPSLGRRRRRODIRHGNPALOC-----GOSOEBAVGLVATWYGTE 688
DB 540 SSCSSTFPT-AKRRTRRDRIKNGDPLTHCSDLQHDHNPSCQTLLEER-----IYGYE 590
QY 689 HNSTFLECLPKSPQAVRWMLORPDEGPDQVYTERVLTENGELFRRLSRDAGTYTC 748
DB 591 NSTFLECLPKSPQAVRWMLORPDEGPDQVYTERVLTENGELFRRLSRDAGTYTC 650
QY 749 TLEHGFSTQVYRLVIVASQDLNLFPPPEKPEEPARGGLASTPP--KANYKDIIQL 806

```

```

Db      651  HAVEHEFMOTLKVLTVLEVIDTHELLEELLKHKDDGDSKTEKMSSSMTPSOKWYRDPMDL 710
Qy      807  IGFANLPRVDEYCEYRWKCRTECCSGCFRSRSRGAQKRGKSNAGLEIGKKMSRVAEHN 866
Db      711  INHPNINTNDEFCQYWKRRDKQ----RRPFGHSQSSNMKMKHMOESKGRNRRTHEFE 766
Qy      867  RPPREV 872
Db      767  RAPRSV 772

RESULT 6
ID      SM3A_MOUSE
AC      008655: 062180: 062215;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Semaphorin 3A precursor (Semaphorin III) (Sema III) (Semaphorin D)
       (Sema D).
GN      SEMA3A OR SEMAD OR SEMD.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
       NCBI_TaxId=10090;
       [1]
RP      R1P      SEQUENCE FROM N.A.
RC      STRAIN=NMRI; TISSUE=Embryo;
RX      MEDLINE=95267431; PubMed=7748561;
RA      Pueschel A.W., Adams R.H., Betz H.;
RT      "Murine semaphorin D/collapsin is a member of a diverse gene family
       and creates domains inhibitory for axonal extension.";
RL      Neuron 14:941-948(1995).
       [2]
RP      R2P      SEQUENCE FROM N.A.
RX      MEDLINE=97470885; PubMed=9331345;
RA      Taniguchi M., Yuasa S., Fujisawa H., Naruse I., Saga S., Mishina
       T.;
RT      "Disruption of semaphorin III/D gene causes severe abnormality in
       peripheral nerve projection.";
RL      Neuron 19:519-530(1997).
       [3]
RP      R3P      SEQUENCE OF 107-772 FROM N.A.
RX      TISSUE=Retal brain;
RA      MEDLINE=95267432; PubMed=7748562;
RT      Messersmith E.K., Leonardo E.D., Shatz C.J., Tessier-Lavigne M.,
       Goodman C.S., Kolodkin A.L.;
RT      "Semaphorin III can function as a selective chemorepellent to pattern
       sensory projections in the spinal cord.";
RL      Neuron 14:949-959(1995).
       -1- FUNCTION: PLAY A ROLE IN GROWTH CONES GUIDANCE. MAY FUNCTION TO
       PATTERN SENSOR PROJECTIONS BY SELECTIVELY REPELLING AXONS THAT
       NORMALLY TERMINATE DORSALLY.
       -1- SUBCELLULAR LOCATION: Secreted.
       -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN EMBRYONIC DEVELOPMENT
       (E11) IN DISTINCT REGIONS OF THE NEUROECTODERM AND MESODERM.
       EXPRESSION BECAME MORE EXTENSIVE AT LATER STAGES.
       -1- DOMAIN: STRONG BINDING TO NEOPHILIN IS MEDIATED BY THE CARBOXY
       THIRD OF THE PROTEIN.
       -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
       -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
       -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
       -----
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation
Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc      or send an email to license@isb-sib.ch).
Cc      EMBL; X85993; CAA59985.1; -
Cc      EMBL; D85028; BAA19773.1; -

```

DR	EMBL: L40484; AAA33934.1; .	
DR	MGI: 107558; Sema3a.	
DR	InterPro: IPR003599; Ig.	
DR	InterPro: IPR003006; Ig_MHC.	
DR	InterPro: IPR003659; PSI.	
DR	InterPro: IPR001627; Sema.	
DR	Pfam: PF00047; Ig_1.	
DR	Pfam: PF01403; Sema; 1.	
DR	SMART; SM00409; Ig; 1.	
DR	SMART; SM00423; PSI; 1.	
KW	Signal; Immunoglobulin domain; Multigene family; Neurogenesis.	
KW	Developmental protein; Glycoprotein.	
FT	SIGNAL	1 20
FT	CHAIN	21 772
FT	DOMAIN	240 538
FT	DOMAIN	643 730
FT	DOMAIN	728 770
FT	DISULFID	650 723
FT	CARBOHYD	53 53
FT	CARBOHYD	125 125
FT	CARBOHYD	591 591
FT	CONFLICT	133 193
FT	CONFLICT	207 207
FT	CONFLICT	253 253
FT	CONFLICT	352 352
FT	CONFLICT	403 403
FT	CONFLICT	571 572
FT	CONFLICT	616 620
FT	CONFLICT	623 623
SO	SEQUENCE	772 AA: 88799 MW: 88908528810AEC3 CRC64: E89A0B528B10AEC3

Query Match 38.8%; Score 1843; DB 1; Length 772;
Best Local Similarity 46.6%; Pred. NO. 9.6e-129;
Matches 366; Conservative 138; Mismatches 248; Indels 34; Gaps 11.

QY	101	ICWAGGILHCGSSGPRSPVRYPLRLSYROLLSANKRAITLFGOGSLNDOAMTIDYR	160
Db	7	IACLEWGLVLLTARAYANANKNNVPRFKJSTKMLSSNNVITTINGLANSSSSYTFLDEER	66
QY	161	DRLFLGGLDALYSLRLDOAMPDRREVLWPRPGOREECYRKGRDPLETCANFVRLQPN	220
Db	67	SRUYAGKDHFESFNLVNI - KDFOKIIVVSYSTRDECKWACKDILKECANIKYLEAN	125
QY	221	RTHLLACGTGAAPORPCALLITVGNRE - HYLHPEBSVESGRCRCHESSRPASTFIDE	279
Db	126	QTHLACGTGAHPRICSTYEVENHEDNIEFKLDQSHFENGRCSPYDKLTASLLIOE	185
QY	280	LYTGLTAOFLIGREAMIFSGGRPALRSDS - DOSLHDPFMAARIPNSOODNDKYF	338
Db	186	LYSGTAADFMRGDAIFPLTIGHNHRIRREQDHSRNLNDRPISAILHIESDPEDDKYF	245
QY	339	FFSEFVSPDGC - GSNHVTVSRVGRVCVNDAGQRYLVNKKSTFLKARLYSVGFGAET	397
Db	246	FFREN - AIDESHGSKATHARIQDICKNDPFGHBRSLVNKKWTFELKARLYSPVGNIGDT	303
QY	398	HFDOLDEVLFLMPKAGSLEYVALFESTSAVQFGRVAVCVUHNADLWVENGFNGFARDDQ	457
Db	304	HFDOLDYFLNASKDPKPIYGVFTTSSNIFKSGAVCMYSMSDVRVFLGAYARHGDG	363
QY	458	HOMGBYGGKVPFPRBGVSRKMTADPGARFEGSTKQPRDEVULOFARANLIMFVPRPRGR	517
Db	364	YOMVYQGRVRYPRRGTCPSKTFG - - - GFDSKDLPRDVLITFASHPRAMVPRPEINNR	419
QY	518	PVLKTHLAQDLHIVDVRVAEDSTYVITLCTDGSGLYKVALIALOAGSAPEEVLAE	577
Db	420	PIMKTIDVNYOFTIIVDVRVAEDQYVMEITGIDVGLTKVSVSPKETJWHDLEVLLE	479
QY	578	LOVEKVPPTIEMETISVKOMLYUGSRGLVQALRHOCETGTACAECCLARDPVCAND	637
Db	480	MTVREPTTISAMELSTKOQOLYIGSTAGVADPLHRCDIYGRKACBECCLADPVCAMD	539
QY	638	ASCTHYRSLGKRRFRRODIRHGNDAIOLC - - - - - GSOEEAVGLVAATVYCTE	688

Db 540 SSCSRFFPT-AKRRTRRQDRIKNDPLTHCSQDQHNDHNGPSLEER-----IIYGE 590

Qy 689 HNSTFLECLPKSPQAAVRLQRPDEGPDYKDERVLHTERGLLRLSLFDAGTYTC 748

Db 591 NSTFLECLPKSPQAAVRLQRPDEGPDYKDERVLHTERGLLRLSLFDAGTYTC 650

Qy 749 TLEHESQTVARLALVYVASOLDLFPPEPKPEEPARGGLASPP--KMYDIDL 806

Db 651 HAVHEFMQTLKLVLEVDTEHLEHLKDDGSGKIKEMSSMTPSQKYVYDFML 710

Qy 807 IGFANPRVDECEWYVCHTTECSGCFRSRSGKQARGKSWAGLELKKMSRVHAENH 866

Db 711 INPNLNTDEFEQYWKDRKO---RRQPGHSGSSNMKMKHMGESKGRNRRTHEE 766

Qy 867 RPREV 872

Db 767 RADRSV 772

RESULT 7

S2LB_BRARE STANDARD; PRT; 778 AA.

ID S2LB_BRARE

AC Q9W686;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Semaphorin 21B precursor (Semaphorin 1B) (Sema-21B).

GN Sema21B OR Sema3AB.

OS Brachydanio rerio (zebrafish) (zebra danio).

OC Brachydanio rerio; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Osteichthyes; Cypriniformes; Cyprinidae; Danio.

OC Cypriniformes; Cyprinidae; Danio.

OX NCBI_TaxID=7955.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99425174; PubMed=10495275;

RA Roos M., Schachner M., Bernhardt R.R.;

RT "zebrafish semaphorin 21b inhibits growing motor axons in vivo.";

RL Mech. Dev. 87:103-117(1999)

CC -1- FUNCTION: MIGHT NORMALLY INFLUENCE THE MIDSEGMENTAL PATHWAY CHOICE OF THE VENTRALLY EXTENDING MOTOR AXONS BY CONTRIBUTING TO A REPULSIVE DOMAIN IN THE POSTERIOR SOMITE.

CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN RHOMBOMERES THREE AND FIVE, AND IN THE POSTERIOR HALF OF NEWLY FORMED SOMITES WHICH IS AVOIDED BY VENTRALLY EXTENDING MOTOR AXONS.

CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL AF083382; AAC28103.1; -

DR ZFIN; ZDB-GENE-991209-6; Sema2ab.

DR InterPro: IPR003599; 19.

DR InterPro: IPR003006; 19_MHC.

DR InterPro: IPR003659; PSI.

DR InterPro: IPR001627; Sema.

DR Pfam: PF00047; Ig_1.

DR Pfam: PF01403; Sema; 1.

DR SMART; SM00409; IG; 1.

DR SMART; SM00423; PSI; 1.

KW Signal; Immunoglobulin domain; Multigene family; Neurogenesis;

KW Developmental protein; glycoprotein.

FT SIGNAL 1 17

FT CHAIN 18 778

FT DOMAIN 241 539

SEMA.

FT DOMAIN 645 723

FT DOMAIN 721 776

FT DISULFID 652 716

FT CARBOHYD 54 54

FT CARBOHYD 127 127

FT CARBOHYD 593 593

SO SEQUENCE 778 AA; 88904 MW; 4D36F433AE21895 CRC04;

Query Match

Best Local Similarity 47.9%; Pred. No. 1,6e-128;

Matches 378; Conservative 123; Mismatches 258; Indels 30; Gaps 11;

38.8%; Score 1840; DB 1; Length 778;

Qy 99 WAICWLLGGLL-HGSSGSPSPGSPVRLSTYDLSANSAFLPOGSLNQAMTYD 157

Db 5 WIVVLLIWTLLAPERGFAVRSKSNVPRKPSYKEMESNMLTFNGIANSAYHTFLD 64

Qy 158 EYRDLPLFGIDALYSRLDQAMPDPRREVLMPPOGOREECVRCGRPLFCANFVQLQ 217

Db 65 EERGLFVGAKDHLSEVLDINDDQSLSNPSPSRDECKNAGKQKCAKAFIVLQ 124

Qy 218 PNRTHLLACGTGAFOPCALITVGRGEHVLHLEPGVESGRCRHEPSRPASTFID 277

Db 125 PENOHLIACGTGAFHPCAHAYEVGKRESEDMTFLGSSFEENGKSPYDKLOTASMLID 184

Qy 278 GELVYGLTADFLGRAMIFRSGPRPALRSDS--DOSLLHDPRTVMAARIIPNSQDNKV 336

Db 185 GELVYAGTSADPFWGRDPAFRTLGKHPHPIKTEBDHDSRWLNDPFEVSHLIPSDNADDKT 244

Qy 337 YFEFSETPVSPDGGG-NHNTVSRGVYCVNDAGGQGVLYNKSTFLKARLYCSVPGPGA 395

Db 245 YAFERRN--AIDGEOISKATTHRIQOLCKNDGGRSLVKNWTFLLKRLAYCVGLNGI 302

Qy 396 ETHEFDLEDVFLIMPRAGKSLVYALFSTVSAVFOGFAVCYHMDIYEVNGPRAHDG 455

Db 303 DTHFDELQVFLIMSKDPKNPITVAVFTSSNIFGSAVCMSMDIRVFLGPAHBDG 362

Qy 456 POHQMPYGYGKVPFRPGCPSPKMTAGPGRGPGSTKDPDEVLOFARAHPLFWFVPRH 515

Db 363 PNYQWVFLNRPVPRPGCPSPKTF---DGEFTKDPDDVITTRAKSHPMYNTVFFIN 418

Qy 516 GRVLYKTHLAQOLQIVDVEAEEDGYVIFLGTSGSVLYKVALOAGSAREEVL 575

Db 419 NHPIIKTIDVDQFOIVDVEAEEDGYVIFLGTSGSVLYKVALOAGSAREEVL 478

Qy 576 EELQVKEVPTITEMEISYKRMLYGSRVQAOLRHOCETTYGTRACACCLARDPYCAW 635

Db 479 EEMTVREPTAITAMELSTKOOLYLGSAIGVOMDLHRCQVYGRACACCLARDPYCAW 538

Qy 636 DGASCHYRSLGKRRFRQDRIKNGDPLTHCSQDQHNDHNGPSLEER-----VGLVATVYGEHNST 692

Db 539 DGSCSRFFPT-AKRRTRRQDRIKNDPLTHCSQDQHNDHNGPSLEER-----VGLVATVYGEHNST 596

Qy 693 FLECLPKSPQAAVRLQRPDEGPDYKDERVLHTERGLLRLSLFDAGTYTC 752

Db 597 FLECLPKSPQAAVRLQRPDEGPDYKDERVLHTERGLLRLSLFDAGTYTC 656

Qy 753 HGSQVYVRLALVYVASOLDLFPPEPKPEEPARGGLASPPKMYDIDL 812

Db 657 HGIQTLRLTLVIPAHLDDLHARDPPTNDPANG-----KMYRDFLSLNPSP 709

Qy 813 PRVDECEWYVCHTTECSGCFRSRSGKQARGKSWAGLELKKMSRVHAENH 863

Db 710 NSVDQCEQWYKRRKOROKANILHSSQSHQILHSSQSHAKWMLDENKGRNRRTHEE 769

Qy 864 EHNTPREV 872

Db 770 EMGRAPRSV 778

RESULT 8

SM3A_HUMAN STANDARD; PRT; 771 AA.

ID SM3A_HUMAN

AC Q14563;

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Semaphorin 3A precursor (Semaphorin III) (Sema III).
 GN SEMA3A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=94094332; PubMed=8269517;
 RA Kolodkin A.L., Matches D.J., Goodman C.S.;
 RT "The semaphorin genes encode a family of transmembrane and secreted
 growth cone guidance molecules.";
 RT Cell 75:1389-1399(1993).
 RN [2]
 RP SEQUENCE OF 1-37 FROM N.A.
 RA Woessner J., Mink P., Hinds K., Strommatt C.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 39-182 FROM N.A.
 RA Rohlfing T., Tin-Mollam A.M., Duckels G.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
 CC CONS. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
 CC BY A MOILITY-INHIBITING MECHANISM. BINDS TO THE COMPLEX
 CC NEUROFILAMEN-1/PLEXIN-1 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- DOMAIN: STRONG BINDING TO NEUROFILAMEN IS MEDIATED BY THE CARBOXY
 CC THIRD OF THE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L26081; AAA65938.1; -;
 DR EMBL: AC004451; AAC06185.1; -;
 DR EMBL: AC004848; AAC78622.1; -;
 DR MIM: 603961; -;
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF00047; Ig_1.
 DR Pfam: PF01403; Sema_1.
 DR SMART: SM00409; Ig_1.
 DR SMART: SM00423; PSI_1.
 DR Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
 DR Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
 DR Developmental protein; Glycoprotein.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 771 SEMAPHORIN 3A.
 FT DOMAIN 240 538 SEMA.
 FT DOMAIN 642 729 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 727 769 ARG/LYS-RICH (BASIC).
 FT DISULFID 649 722 BY SIMILARITY.
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 771 AA: 88689 MW: 9985f8d3baed8456 GRC64;
 Query Match 38.6%; Score 1830; DB 1; Length 771;
 Best Local Similarity 46.5%; Pred. No. 8,8e-128;
 Matches 366; Conservative 139; Mismatches 244; Indels 38; Gaps 13;

QY 101 ICWLLGILLHGGSSGSPSPBRLRLSTYRDDLSANRSATFLGPGQSLNLAQWLYDEXR 160
 DB 8 VC-LFWGVLLTARANYQNGKNVPRKLSTYKEMLESNVITFNGLANSSSYITFLDEER 66
 QY 161 DRLFGGLDALYSRLDQAWPDPREVLPPOPQAECEVCRKGRDPLTCANFVRLQPHN 220
 DB 67 SRLVGAADHIFLFDLVNI-KDQKIVPVSTYRDECKKMGKDLKCANFIKVLKAYN 125
 QY 221 RTHLLACGTGAFOPTICALITVGNGE-HVLLLEPVSYSGRCPCHPSRPFASFIDGE 279
 DB 126 QTHLYACSTGAFFHICITYIEIGHPEDEIFLKNHFNENGKSKSYDPKLLTASLLDGE 185
 QY 280 LYTGTADFLREAMIFRSGGPRALRSDS-DQSLHDPRFVMAARIPENDODNDKYF 338
 DB 186 LYSGADEWGRDAIFETLGHHPHIRTQDHSRWLNDPKFISALISESDNPEDKYF 245
 QY 339 FFESEVSPDG-GSNHVVSVRGVGVADGAGRVLVNKMSTFLKARLYCVSPGCAET 397
 DB 246 FFEREN-AIDESHGKATHARIGQICKNDGHSVLNKMSTFLKARLYCVSPGNGIDT 303
 QY 398 HFQDLQEDVFLMPKAGSLLEVYALFSTVSAVGGFACVYHMAIDWIEFNGPFARHDPQ 457
 DB 304 HFEDLOVFLNFKDPKPNPVYGVFTSSNIFKGSACVMSMSDVRVFLGPAHRRGN 363
 QY 458 HONGPYGKVPFPPGPGPSKMTQPGRPSTQDYDEVLOFRAHPLMFWPRPHGR 517
 DB 364 YQWVPYGGVPPPPGPGPSKMTG-----GFDSTKDLDPDVTTFARSHAPVNPVFPNNR 419
 QY 518 PVLKTHLAQQLQIVDVYEAEDGTIVIFLGTDSGSYLKVIALQGSAPPEVLEE 577
 DB 420 PIVKTVNQQVQIVDVYDAEDGQYVMEIGTVGTVLKYVISPREMVDLEEVLEE 479
 QY 578 LOYEKVPPTTEMEISKRMQLVYSGRLVGAQLRHOCETTYGACACCLARDPYCAMDG 637
 DB 480 MYFREFPTAISAMELSTKQQLITGSTAGVADLPRLHRCIDYGAACBCCCLARDPYCAMDG 539
 QY 638 ASCTHKRPSLGRFRFRDRIHGNPALOCL-----GQSEEAAGLVAAATVNGTEH 669
 DB 540 SACSRYEPT-AKRRTRRQDRIHGNPALOCL-----GQSEEAAGLVAAATVNGTEH 669
 QY 690 NSFFLECLPSQAAVRLQRPDGPDOYKDERVLTERRLSRFAGTYTCT 749
 DB 591 SSTFLCSPKSQALVYWCQRRNERKEIRVDHIIIRDOGLLRLSLQKSCGNVLC 650
 QY 750 TLEHGFQTVVRLALVYVASQDNLFPPEKPEDEPARGLASTPP--KAWKDLQQL 807
 DB 651 AVDHGFIQTLKTVLEVIDEHLHLDHDDGSGKTKMSMSMTPSQKWTYRDMQL 710
 QY 808 GFANLPRVDYECRWVCRGTTECGSGCFRSRSRGQAQKRS--WAGLELKKMKSRVAEH 865
 DB 711 NHENLMTDEFCBQVWKRDRKQ-----RRPRGHTPGNSNMKKHLQENKKGNRRTHFE 764
 QY 866 NRTPREV 872
 DB 765 ERAPRSV 771
 RESULT 9
 ID SM3D_CHICK STANDARD: PRT: 761 AA.
 AC 090663;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Semaphorin 3A precursor (Collapsin-2) (COLL-2).
 GN SEMA3D OR COLL2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 ON NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.

CC	EMBL; U28240; AAA86896.1; -.	
DR	InterPro; IPR003006; IG_MHC.	
DR	InterPro; IPR003600; IG_Like.	
DR	InterPro; IPR003659; PSI.	
DR	InterPro; IPR001627; Sema.	
DR	Pfam; PF00047; Ig_1.	
DR	Pfam; PF01403; Sema; 1.	
DR	SMART; SM00410; IG_Like; 1.	
DR	SMART; SM00423; PSI; 1.	
KW	Signal; Immunoglobulin domain; Multigene family; Neurogenesis;	
KW	Developmental protein; Glycoprotein.	
FT	SIGNAL	1 24
FT	CHAIN	25 761
FT	DOMAIN	245 543
FT	DOMAIN	646 726
FT	DOMAIN	727 757
FT	DISULFID	653 719
FT	CARBOHYD	127 127
FT	CARBOHYD	595 595
SQ	SEQUENCE	761 AA: 87300 MW: 3E09AE3DBA53F46B CRC64;

[illegible]

Db	298	CS1PBGPGADJTHDELODJFLJLSTROBERNDLVJGVFTTSSVFKGSAVCVSMADIRAVE	357
Qy	447	NGPFAHNDGEOHOMGPRYGKVPFPPRGVCSKMTAOPGRFGSTKXDPDEVLQOFARHPL	506
Db	358	NGPFAHNSADHWRVQYEGRIPIYPRPCTPSK--TYDP--LIKSTPFPDEVLSEFKRPL	414
Qy	507	MEWVYPRHRGPRVYLVKTHLAQHLQIYVDHVEVAEDGTVDFIPLGTSGSVLYKVALQAG	566
Db	415	MYKSYVPLTGGPRVTRINVDYRLTLQIYVDHVAEMDDGDVDYFIPLGTIDIGVLYKAVSITKEK	474
Qy	567	SAEBEYVLEELQYFKVPTPIITEMEISVKQOMLYVSRGLVQAOLRLHOCSTYGTAAECC	626
Db	475	WTK--EEVLELELQIFKHPHSFISTMEISOKOQOOLYIGSRDLVLSLRCHTYKACADCC	533
Qy	627	LARPYCAMOASCTHNPSPISGKRFRPRRODIRHGNPMLQCLQSGQSEEAAGLVAA--TWY	684
Db	534	LARDPYCAMDNSCSTRAPF--SKRARQDYKYDDPVAQCM---DVEDDLSHNEADEKVI	589
Qy	685	YGTENHSTFLEELCKRSPQAAVFWLLOPRGDEGPDQVTKDERVLTETGGLFRLISRPDAG	744
Db	590	FGIENHSTFELCIKPSQASIRWYIQSGHEHREBELADERIKITEHGLLIRSLQORDAG	649
Qy	745	TYTCTLEHGSQVYRLALVVIASOLDNLFPEPPEPPEPARGGLASIPPKMYKDIL	804
Db	650	AYFCAQGHETHTHTVKLNTLVNENGQNES---TQKTEDEGVNRLDLESRLRYKDYI	705
Qy	805	OLIGFANPRVDECEYRWCRGTTECSGCFSPRSRGQOARGKSMAGLELCKMKSRVAE	864
Db	706	OLVSSPSF-SLDEYCEQMW-----HKKRRORNRKGAKKWHQEMKKKRRRHE	754
Qy	865	HNKRP 869	
Db	755	PARPP 759	
RESULT 10			
ID	SE1A_BRARE	STANDARD:	PRT: 860 AA.
AC	Q9WJ71:		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Semaphorin z1a precursor (Semaphorin 1A) (Sema-z1a).		
GN	SEMAZ1A OR SEMA3A.		
OS	Brachydanio rerio (Zebrafish) (Zebra danio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;		
OC	Cypriniformes; Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo.		
RA	MEDLINE=99313409; PubMed=10386638;		
RA	Yee C.S., Chandrasekhar A., Halloran M.C., Shoji W., Warren J.T.,		
RA	Kuwada J.Y.;		
RT	*Molecular cloning, expression, and activity of zebrafish semaphorin		
RT	z1a.";		
RL	Brain Res. Bull. 48:581-593(1999)		
CC	- FUNCTION: MAY INFLUENCE OUTGROWTH BY A VARIETY OF GROWTH CONES		
CC	INCLUDING THOSE OF THE POSTERIOR LATERAL LINE GANGLION.		
CC	- SUBCELLULAR LOCATION: Secreted (By similarity).		
CC	- DEVELOPMENTAL STAGE: EXPRESSED IN HIGHLY SPECIFIC PATTERNS WITHIN		
CC	THE DEVELOPING EMBRYO.		
CC	- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.		
CC	- SIMILARITY: CONTAINS 1 SEMA DOMAIN.		
CC	- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		

```

CC or send an email to license@lsb-sib.ch).
CC EMBL: AF086761; AADA3964.1;
DR ZFIN: ZDB-GENE-991209-3; sema3aa.
DR InterPro: IPR003599; 1g_1MHC.
DR InterPro: IPR003006; 1g_1MHC.
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF00047; 1g_1.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00409; IG; 1.
DR SMART: SM00423; PSI; 1.
KW Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1 17
FT CHAIN 18 860
FT DOMAIN 241 539
FT DOMAIN 645 724
FT DOMAIN 722 858
FT DISULFID 652 717
FT CARBOHYD 53 53
FT CARBOHYD 126 126
FT CARBOHYD 593 593
SQ SEQUENCE 860 AA; 97263 MW; 5FD4C12194F5165C CRC64;

Query Match 37.3%; Score 1770; DB 1; Length 860;
Best Local Similarity 44.9%; Pred. No. 2.8e-123;
Matches 347; Conservative 144; Mismatches 245; Indels 36; Gaps 10;

OY 104 LGLGLHGGSSGSPSPVRLRLSLYRDLISANRSLFLCPOGLNIQMYLDEYRDL 163
DB 10 LCCVALPGRVAPQHTENPVRLSLYRDLISANRSLFLCPOGLNIQMYLDEYRDL 69
OY 164 FLGLDALYSLRLDQAMPREVLPPQPGORECVKRGKRDPLECANFVVLDPHNHT 223
DB 70 LVGEDVEFSEFDLYNINRDYQQLAMPATPSKRDCKWAKNDRDCSNFVAVLSYQTH 129
OY 224 LMGCGTAPQTCALLIVGRGE-HVHLEPGSVESGRCPHEPSPRPAFTIDELYT 282
DB 130 IYCGTGAFPHICSELMGKRAEDNIFRLDANYFENGRGKSPYDKKQSSILLDGEYS 189
OY 283 GLTADFLGRAMIFRSQGPRLALRSDS-DQSLDHPREVAARIPENSDDNDKVFEEFS 341
DB 190 GTSADFGRGFAIFRLGSHHPIRTEDHDSRWLNEPFLGHLIPESDNEPDDKIFLFEK 249
OY 342 EYTPSPDG-GSNHYTVSRVGVCVNDAGGRVLYNKKWSTFLKARLYCSVPGGAEYTHFD 400
DB 250 EN--AMDGHTGKATISRIQOLCKNDMGHRSIVNKKWTFLKAKLTGSPGLNGIDTHFD 307
OY 401 QLEDFELMPKAKSLLEVVALFSTVSAVFGGFVAVCYHMADINEVFNPRRAHNDGPOHOW 460
DB 308 ELQDVLMSAKDKRNKYITAVFTSSNIFGSAICMSMADIRVFLGPRVHNRGPNYQW 367
OY 461 GRYGKVPFRPVCPSCSKMTAOPGRPFSTKDYRDEVLQARAHPLMFVPRRHGRPV 520
DB 368 VPEQGHVPRPFCPSKPTG---GFDSTKDLRDPDYTFARLHPANYNVQVPGKFPV 423
OY 521 VKTHLAQQLQIVVDVEADGTYVIFLGCTDSGVLYKVTALQAGSAPEVEVLELOY 580
DB 424 VRTNVEYQYQLQIVVDVEADGTYVIFLGCTDSGVLYKVTALQAGSAPEVEVLELOY 483
OY 581 FVVPPTFMEISVKRQMLYVGSRLGVAOLRLHOCGYGACACCLARDPYCAMPGASC 640
DB 484 PREPPTITMELSTKQOQLYLSGLDLSIQMPLHRCVYGAACECLARDPYCAMPGASC 543
OY 641 TVYRSLGRARRRQDIRGNPALQC-LGQSQEEAVGLVATATWYTGTEHNSTFELCP 698
DB 544 SYTFPT-ARRRTRRQDIRGNPALQC-LGQSQEEAVGLVATATWYTGTEHNSTFELCP 602
OY 699 KSPQAAVRLRLQPGEGEDQYKTBERVLTHERGLFRLSRPDAGTYCTTLEHFSQT 758
DB 603 KSRALLITWQLOKPNDEKHEIVIDERLSLTQGLLIRSLTQADSQVFLCHAVEHGFIDP 662

```

```

OY 759 VRLALVIVASQDLNLEPPEKPEEPARGGLASTPPKAWKIDILQIGFANLPRVDEY 818
DB 663 LRRINLOVIPSQVSELLLRAGINDKPA-----PKHLMWRDFPSLLEHNDLVNDEF 716
OY 819 CRRVNCRGTTESGCRSRKQKARGKSWAGLELCKKRSVYHAEHNTPR 870
DB 717 CERIM-----KREKKRPGKRAKRVNPG---TGVSIKNEKTPQ 750

RESULT 11
SMZ2_BRARE
ID SMZ2_BRARE STANDARD; PRT; 764 AA.
AC Q9WGG6;
DI 30-MAR-2000 (Rel. 39, Created)
DT 30-MAR-2000 (Rel. 39, Last sequence update)
DE 01-MAY-2002 (Rel. 41, Last annotation update)
DE Semaphorin 22 precursor (Semaphorin 2) (Sema-22).
GN SEMA22 OR SEMA2.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99112778; PubMed=9915572;
RA Halloran M.C., Severance S.M., Yee C.S., Gemza D.L., Raper J.A.,
RA Kuwada J.Y.;
RT "Analysis of a zebrafish semaphorin reveals potential functions in
RT vivo."
RL Dev. Dyn. 214:13-25(1999).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE GUIDANCE OF SEVERAL AXON
CC PATHWAYS.
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN A DYNAMIC AND RESTRICTED PATTERN
CC DURING THE PERIOD OF AXON OUTGROWTH.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC
DR EMBL: AF124485; AAD21310.1; -.
DR ZFIN: ZDB-GENE-990715-2; sema2.
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00423; PSI; 1.
KW Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1 41
FT CHAIN 42 764
FT DOMAIN 261 559
FT DOMAIN 661 740
FT DOMAIN 741 762
FT DISULFID 668 733
FT CARBOHYD 143 143
FT CARBOHYD 490 490
FT CARBOHYD 610 610
SQ SEQUENCE 764 AA; 87859 MW; A3ED95C2C479D7AE CRC64;

Query Match 37.2%; Score 1766.5; DB 1; Length 764;
Best Local Similarity 45.3%; Pred. No. 4.4e-123;
Matches 359; Conservative 137; Mismatches 245; Indels 52; Gaps 18;

OY 84 RRQRCQFSPMSAPALWILGGLLHGGSSGSPSPVRLRLSLYRDLISANRSLFL 143
DB 111

```

```

Db 10 RRORQVYRGFRSCAM---WSTSYMLFFSLPEGNCKMSESLPRYKLGKYLHRSVYFT 66
QY 144 GPOGSLNLOAMYDEYDRFLFLGGLDALYSRLDOAMPDREVIMPPQPGQRECVKGR 203
Db 67 GSSEGHFQVLLDEEBSRLLGAKHVVLLDPDINIKRHKLSWSPASRDREVEMCILAGK 126
QY 204 DPLTECANFVYVLOPHNRTHLLACGTAFOPTCALITY-GHRGHVHLLEFGSVESGR 262
Db 127 NPLTECANFVYVLOPHNRTHLLACGTAFOPTCALITY-GHRGHVHLLEFGSVESGR 186
QY 263 CPHEPSRPFASSTFIDELTYGLTADFLGREAMIFRSGGPRP---ALNSD-SDQSLHDP 318
Db 187 CFPDPPQPFASVLTLDQYLYAGTASDFLGKSTFRSLCPRPHQOYRTIDISEDYINEGK 246
QY 319 FVMAARIPENSDDNKVYFFEFSTVSPDGSNHYV-SVGVVCVNDAGGVVIVNKK 377
Db 247 FISHPISDYNPDNDKTYEFFRFA--SRDSTDKSVLSHVARICRDVAGLSLNK 304
QY 378 STFKARLVCSVPGGAETFFDLEDFLLMPKAKSLLEYALFSTVSAVFOGFAVCY 437
Db 305 TTFKALVCSIPGPDVDTHFDELQDIFLLPSRDEKNPMYGVFTTSSIFKSAVCY 364
QY 438 HMADIWEFNGPFAHRDGPQHQWGPYGGKVPFPPGVCPSKMTAQPGRFSTGYDPEV 497
Db 365 TMEDIRAFNCPYAHKRGDPHWRWEYEGRIPIYPRGTCPSR-TYDP--HIKTKDFPDEV 421
QY 498 LQFARHPMWPVPRPRGRVLYKTHLAOLHIVDRAEDGTIVIFLGDSSVL 557
Db 422 ISFRLHPLMOSVHAPMTGRIFTRINTEYRLTOIVRAVEDQYVMEGLTDMGVL 481
QY 558 KVAL-QAGSGAEPEEVLYEQLQVFKVPTITEMETSVKROMLYVSLGVAOLHOC 616
Db 482 KVSITQENMS--EITLELOQFKNPSPLINMEVSKOOLPFVGSGLVQVSLHNC 539
QY 617 TYGACACCLARDPYCAMDCACTHYRPSLGRKRRRQDRIHGNPALQC-----LQDS 670
Db 540 IYGGGACACCLARDPYCAMDCTGSRYP--SKRRARODIKHGPPSSCHMDTEVDLNR 598
QY 671 QEEENVGLVAVMYGTEHNSFTLECLKSPQAAVRLDORPGDGPPOVKDEVLNTE 730
Db 599 VEER-----VLGVSNSSFTLVCVSKSOALIRWYLVKRGVHRDEIKDERVLTD 650
QY 731 RGLLFRRLSRPDAGTYCTTLEHGFSTQTVRLAVVIVASOLDLFPPEKPEPPANG 790
Db 651 RGLLRWLGROGAGSYFCTSGHREFTRLHLYSLHLDGQINAHOPAIRESNP---- 706
QY 791 LASTPRAKMYKDILQIGFANLPVDEYCEWCGTETGSCCFSRSGKARKKSMAG 850
Db 707 AVTEPRORYKDYLRMLS-GPARSLDEYCEYTW-----HREKKOKRGR-WKH 751
QY 851 L-ELGKKMKSRVH 862
Db 752 VQELKRSRRNRHH 764

```

```

RT "Human semaphorin A(V) and IV reside in the 3p21.3 small cell lung
RT cancer deletion region and demonstrate distinct expression patterns.",
RT Proc. Natl. Acad. Sci. U.S.A. 93:4120-4125(1996).
RN [2]
RA Dente M., Wamsley P.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
CC SPECIFIC TERRITORIES INACCESSIBLE FOR GROWING AXONS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY). ACCUMULATES IN THE
CC ENDOPLASTIC RETICULUM.
CC -1- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY BUT DIFFERENTIALLY IN A
CC VARIETY OF NEURAL AND NONNEURAL TISSUES.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28369; AAD09138.1; -
DR EMBL: U73167; AAC02731.1; -
DR MIM: 601281; -
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF00447; Ig; 1.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00409; Ig; 1.
DR SMART: SM00423; PSI; 1.
KW Signal, Immunoglobulin domain; Multigene family; Glycoprotein.
FT SIGNAL 1 24
FT CHAIN 25 749
FT DOMAIN 239 537
FT DOMAIN 637 717
FT DOMAIN 699 702
FT DOMAIN 724 744
FT DISULFD 644 710
FT CARBOHYD 82 82
FT CARBOHYD 124 124
FT CARBOHYD 427 427
FT CONFLICT 29 36
FT SEQUENCE 749 AA; 83121 MW; 1F3B8F63F59444F3 CRC64;
SQ
Query Match 36.8%; Score 1747; DB 1; Length 749;
Best Local Similarity 47.1%; Pred. No. 1.2e-121;
Matches 367; Conservative 107; Mismatches 239; Indels 66; Gaps 17;
QY 83 GRRORQVYRGFRSCAM---WSTSYMLFFSLPEGNCKMSESLPRYKLGKYLHRSVYFT 66
Db 2 GRAGAAVIFGLA-LTMAY-----GL---GSAAPSP---PRLRLSTFQELQAHGLOTF 47
QY 135 SANRSALIFLPGGSLNLOAMYDEYDRFLFLGGLDALYSRLDOAMPDREVIMPPQPGQ 194
Db 48 SLERTCY-----QALLVDEERGRLFVGAENNVASLINDNISKRKKLAMPVPEV 98
QY 195 REECVARGDPLTECANFVYVLOPHNRTHLLACGTAFOPTCALITYGHRGHV-VLHLEP 253
Db 99 REECVARGDPLTECANFVYVLOPHNRTHLLACGTAFOPTCALITYGHRGHV-VLHLEP 158
QY 254 GSVEGRGRCPRHPSRPFSTFIDELTYGLTADFLGREAMIFRSGGPRPALS-DQGS 312
Db 159 GRIDEKGKSPYDPRHARSVIVGEELYSVADLWGRDFTIRSLGQPSLTERPHDSK 218
QY 313 LLDPRFVMAARIPENSDDNKVYFFEFSTVSPDGSNHYV-SVGVVCVNDAGGVVIVNKK 372

```

```

Db 219 WLNPKVKVFWPDESENDDKITYFEFFRETAVEAARALGLSVSRVGOICRNVDVGORS 278
QY 373 LVNWKSTFLKARLVCSVPGAGAEHFDOLBDVFLMLPKACKSLVLAFFSTVASVPOGF 432
Db 279 LVNWKSTFLKARLVCSVPGAGAEHFDOLBDVFLMLPKACKSLVLAFFSTVASVPOGF 337
QY 433 AVCVYHMADTWENFNGPRARHDGPOHONGRYGKVPFRPGVCSCKMTADGRPGSTKD 492
Db 338 AVCVYHMADTWENFNGPRARHDGPOHONGRYGKVPFRPGVCSCKMTADGRPGSTKD 393
QY 493 YPDEVLOFARHAPLMFMPVRRHGRPVLYKTHLAQOLQIVDVVEAEDGTVDYVFLGTD 552
Db 394 FPDVDVLOFARHAPLMFMPVRRHGRPVLYKTHLAQOLQIVDVVEAEDGTVDYVFLGTD 453
QY 553 SGVSLKVLVIALOGASAPBEVEVLEELQVFKVPTTEMEISVKRQMLVYSGRLVGAQRL 612
Db 454 VGVFLKVLVYISVPGSRPSAEGELLEELHVEDSAVTSQMSKSHQVLVARSANAQIAL 513
QY 613 HOCETTYGTACAECLADPYCAMDGASCTHYRPSLGRFRFRDRIHGNPALQCGSQE 672
Db 514 HRCAGHAGVCTECCCLADPYCAMDGAVACTRPOPS-AKRFRFRDRIHGNPALQCGSQE 572
QY 673 EEAVALGAAATVMTVGTENHSPFLECLTPKSPQAAVRLLORPEDGPDVKTDERVLIHNRG 732
Db 573 P-----ALLEHKVFEVSSAFLECEPSSLOARVMTFORAGVTAHTOYLAERTERTARG 628
QY 733 LLEFRLSRPDAGTYCTTLEHFSQTVRLALVYVASOLDNLF-PEPKPEEPARGGL 791
Db 629 LLEFRLSRPDAGTYCTTLEHFSQTVRLALVYVASOLDNLF-PEPKPEEPARGGL 684
QY 792 ASPPKRWYDILOLI-----GFANLPRVDEYCEYRWCRTGTESSGC-FRFRSGQAR 844
Db 685 -----PKMYDFLOLVPEGGGGSANSLRM-----CRFOPALQSLPLESRGRKRRNR 731

RESULT 13
SM3D_HUMAN STANDARD; PRT; 777 AA.
AC 095025;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 3D precursor.
GN SEMA3D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Mead K., Graves T., Wilson C.;
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INDICES THE COLLAPE AND PARALYSIS OF NEURONAL GROWTH
CC CONNS. COULD POTENTIALLY ACT AS REPULSIVE CUES TOWARD SPECIFIC
CC NEURONAL POPULATIONS. BINDS TO NEUROFILIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- DOMAIN: STRONG BINDING TO NEUROFILIN IS MEDIATED BY THE CARBOXY-
CC THIRD OF THE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AC004957; AAC83081.1; ALT_SED.
CC DR InterPro: IPR003589; Ig.
CC DR InterPro: IPR003006; Ig_MHC.

```

```

DR InterPro: IPR003589; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00409; Ig; 1.
DR SMART: SM00423; PSI; 1.
KW Signal; Immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1..39
FT CHAIN 40..777
FT DOMAIN 257..555
FT DOMAIN 558..738
FT DOMAIN 739..773
FT DISULFID 665..731
FT CARBOHYD 139..139
FT CARBOHYD 607..607
FT CARBOHYD 724..724
SQ SEQUENCE 777 AA; 89651 MW; 3F7B0D7AF50F53BD CRC64;

Query Match 36.3%; Score 1721; DB 1; Length 777;
Best Local Similarity 45.1%; Pred. No. 1e-119;
Matches 346; Conservative 144; Mismatches 229; Indels 48; Gaps 17;

QY 122 SVPLRLSYRDLNLSANRSALFLGPGSLINQAMYLDEYRDLFLGLDALYSLRLDQAMP 181
Db 41 NIPRLKLYKDLILNSNCIPFLGSSEGLDFQTLLEDEGRGLLGANDHIFLFLSVLNLK 100
QY 182 DPREVLPPOGORECEYRKGRDPLTECANFVRYLOPHNRHLLACSGAGAPICALITY 241
Db 101 NFKTIYPAKEREYELCKLACKDANTECANFIRVLOPHNRHLLACSGAGAPICALITY 160
QY 242 G-HGEVHLLEPGSVESGRCRCEPSPASTFIDELTYGLTADFLGREAMIFRSGG 300
Db 161 GYKEDILIFKLDTHNLBSGRKCEPDPOQPAVSMTDEVILSGASDPLGDTAFTSLG 220
QY 301 P---RPLRSD-SQSLIHDPREFYMAARIPENSQDNDKYFFSEFVPPSDGSNHYV 356
Db 221 PTHDHYIRTDISHYWLNGAKKFTGFEIPDYNPDDKITYFFRES--SQEGSTSDKI 278
QY 357 -SRVGRVCMNDAGGQRYLVNWKSTFLKARLVCSVPGAGAEHFDOLBDVFLMLPKAGKS 415
Db 279 LSRGRCKNDYVGQGRSLINKMTFLKARLVCSVPGAGAEHFDOLBDVFLMLPKAGKS 338
QY 416 LEVYALFTSVSAVFOGFAVCYHMDIWEVFNFGFAHRDGPQHONGRYGKVPFRPGVC 475
Db 339 PVYGVFTTSSIFKGSACVYVSMADIRAVFNGPYAKHESADHWRVQYDGRIRPYRGTC 398
QY 476 PSKWTAGRPFPGSTKDYPRDEVLOFARHAPLMFMPVRRHGRPVLYKTHLAQOLQIVD 535
Db 399 PSK-TYPD--LITSTRDPDPDVLSFIKSHSVKSVYPVAGGPFKRINVDYRLQITVD 455
QY 536 RVEAEDGTVDYVFLGTDGSLVLYALOGASAP---EEVLEELQVFKVPTTEMEI 592
Db 456 HVAIEDQYVWFLGTDIGVLYKVS-----SKEKMMEEVLEELQIFKSHSILNML 511
QY 593 SVKROMLYVSRGLVADRLHOCETTYGTACAECLADPYCAMDGASCTHYRPSLGRFR 652
Db 512 SLKQOQLYISRGLVQSLHRCDYTGKACDCLADPYCAMDGASCTHYRPSLGRFR 570
QY 653 RRODIRGNALOC-----LGOSEEAVALGAAATVMTVGTENHSPFLECLTPKSPQAAV 706
Db 571 RRODVAKGDEPITQOCWDLSDSHSHETADEK-----IFGIERNSTFLCEIPSSQATIK 623
QY 707 WLLORPDEGPDVKTDERVLIHNRGLFLRRLSRFDAGTYCTTLEHFSQTVRLALV 766
Db 624 WYIORSGEDEHEELKPERILITEYGLLIRSLQKDSGMYCAKAQEHFTIHTIKVLNV 683
QY 767 IYASOLDNLFPEPKPEEPARGGLASTPRAKMYKQLDILGANLPRVDEYCEYRWCRC 826
Db 684 IENQOMEN--TORAEHEEGVKDLAES--RLRKYDIQILSSPNF-SLDQYCEOMW--- 735
QY 827 TTECSGCFRSRSGKOARGKSMAGL-ELGKKMSRYVAEINRTPREV 872

```


RP SEQUENCE FROM N.A., AND VARIANT MET-503.
 RC TISSUE=Fetal brain;
 RX MEDLINE=96230324; PubMed=8786119;
 RA Xiang R.-H., Hensel C.H., Garcia D.K., Carlson H.C., Kok K.,
 RA Daly M.C., Kerbacher K., van den Berg A., Veldhuis P., Buys C.H.C.M.,
 RA Naylor S.L.,
 RT "Isolation of the human semaphorin III/F gene (SEMA3F) at chromosome
 RT 3p21, a region deleted in lung cancer.";
 RL Genomics 32:39-48(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Nelson J., Biewald T.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 394-436 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=96210603; PubMed=8633026;
 RA Sekido Y., Bader S., Latif F., Chen J.-Y., Duh F.-M., Wei M.-H.,
 RA Albanesi J.P., Lee C.-C., Lerman M.I., Minna J.D.,
 RA "Human semaphorins A(V) and IV reside in the 3p21.3 small cell lung
 RA cancer deletion region and demonstrate distinct expression patterns.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4120-4125(1996).
 CC -1- FUNCTION: MAY PLAY A ROLE IN CELL MOTILITY AND CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity)
 CC -1- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY BUT DIFFERENTIALLY IN A
 CC VARIETY OF NEURAL AND NONNEURAL TISSUES. THERE IS HIGH EXPRESSION
 CC IN MAMMARY GLAND, KIDNEY, FETAL BRAIN, AND LUNG AND LOWER
 CC EXPRESSION IN HEART AND LIVER.
 CC -1- DEVELOPMENTAL STAGE: DETECTED AS EARLY AS EMBRYONIC DAY 10.
 CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in, no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 DR EMBL: U33920; AAC50568.1; -;
 DR EMBL: U38276; AAB18276.1; -;
 DR EMBL: AC000063; AAB46344.1; -;
 DR EMBL: U32171; AAB06011.1; -;
 DR EMBL: U32172; AAB06012.1; -;
 DR MIM: 601124; -;
 DR InterPro: IPR003599; Ig_
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF00047; Ig_1.
 DR Pfam: PF01403; Sema_1.
 DR SMART: SM00409; Ig_1.
 DR SMART: SM00423; PSI_1.
 DR Signal: Immunoglobulin domain; Multigene family; Glycoprotein;
 KW Polymorphism.
 FT SIGNAL 1 18
 FT CHAIN 19 785
 FT DOMAIN 272 569
 FT DOMAIN 671 753
 FT DOMAIN 758 779
 FT DISULFID 678 746
 FT CARBOHYD 53 53
 FT CARBOHYD 126 126
 FT VARIANT 474 474
 FT
 FT VARIANT 503 503
 FT
 FT CONFLICT 153 183 /FTID=VAR_008855.
 FT CONFLICT 270 270 MISSING (IN REF. 2).
 FT CONFLICT 473 473 MISSING (IN REF. 2).
 FT SEQUENCE 785 AA; 88381 MW; FE3FC796EBC1608E CRC64;

Query Match 34.8%; Score 1652; DB 1; Length 785;
 Best Local Similarity 43.0%; Pred. No. 1.4e-114;
 Matches 345; Conservative 128; Mismatches 258; Indels 71; Gaps 17;
 104 LLGGLLHGG-SSGSPFC--PSVPLRLSYRDLSSANSAIFLPGOSLNLQAMYLDEYR 160
 7 LLMASLLTGAMPSPPTODHLPATPRVRLSFKEKATGTAHFNFLLNTDTYRILKKDDH 66
 161 DRLFLGIDLALSLRLQAMPDPREVLMPPPOGORECVKGRPLTECANFVYLOPHN 220
 67 DRMYVSGKDYLSLDLHDINREPLIIMASPOKRIECLVSGKDVNCGNFVRLDPWN 126
 221 RTHLLACGTGAFPTCALITVGHNG-----EHV 248
 127 RTHLYVGTGAYNPMCTYVNGRRAQATPWTQGVAVGRSGRATDGLRPMPTAPRODYI 186
 249 LHLPEVSGSGRGRCRPEPSRPFSTFIDGELTYGLTADFLGREAMIFRSGGPRALRS 308
 187 FYLEPERLESGKCPYDPKLDTRASALINELVAGVYIDEMGTDAIFRTLGKOTAMRT 246
 309 SDOS-LLHDFEFVMAARIPENSDQNDKVFEEFSEVPSPDGSNHTVSRGRVCYND 367
 247 QYNSRWLNDSFTHAELIPDSAEKNDKLFYFRER--SAAEQSPAVYVARIGRICLND 304
 368 GGRVLYNKKSTLTKARLVCSVPGCAETHPDQLEDVFLMKRASKLEVALFTVSA 427
 305 GGHCCLVKKSTFLKARLVCSVPGCAETHPDQLEDVFLMKRASKLEVALFTVSA 364
 428 VFGAVCVYVHMADIWEVNGPRANRDPQHOHNGPVGKVPFRPGVCPKMTAGGRPF 487
 365 VFRGSAVCYVSMADIMVFNFRPAHKEGPNYCMPPSGKMPYRPGTCPEG-TFTPS-M 421
 488 GSKVDPEVLOPARANPLFMFVPRHGRVLYKTHLAQQLHQLVVDREYEDGYDYI 547
 422 KSKVDPEVLEINEMRSHPLMYQAVYPLQRRPLVFRGAPRLITTVADVDADGVEVL 481
 548 FLGTDGSLVLYALQAGSAPEEYVLELQYKPTPTTEMEISVKQMLYSGRLCV 607
 482 FLGTDGTYQKIVLPK-DQELELELLEVEYFKDPAVKVWTJTSKKQOLYVASAVG 540
 608 AOLRLHOCETVGTACAECLARDPYCAMDGASCTHYRPSLGRFRFRDRIHGNPALCL 667
 541 TRLSHRCQAYGAACADCLARDPYCAMDGQACSRYTAS-SKRRSRQVRRHGNPLRQCR 599
 668 G--QSOEEVAVGLVAATMYVGTENHSTFLECLPKSPQAAVRLDOR-PGDEGPDVYKIDE 724
 600 GPNANANKMAV-----ESVQYGVAGSAAFLECPRSPQATVVKWLFQDRDGR-RREIRAD 654
 725 RVLHTEGILFRRLSRFAGTYTCTTLEHGSQTVVRLALVTVASOLD-NLEPPEKPE 783
 655 RLRTREGILLRLALQDSRGIVSCATENNFKHYVTRQVLANHGDVAHVALFPLLSA 714
 784 EPPARGGLASTPRKAWYDIQILGFANLPRVDEYCEYVWCR--GTTECGSCFRSRSG 840
 715 PPPGAG---PPLPPYOLAQLAQPVEGLIHQYCGYVWRHVPSPREPARGAPSPBPQ 770
 841 KQARCKSWAGLELGRKMSRYH 862
 771 DQ-----KKPRNRH 780

Search completed: October 9, 2002, 14:22:39
 Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 13:17:28 : Search time 68 seconds
(without alignments)
2226.038 Million cell updates/sec

Title: US-09-813-290-2
Perfect score: 4746
Sequence: 1 MACALAGKVFPGSWPVMHK.....KKKSRVHAENRTPREAVT 875

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 17294929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	4215	88.8	782 4 Q9NS98	Q9NS98 homo sapien
2	3913	82.4	725 4 Q9H703	Q9H703 homo sapien
3	2052.5	43.2	775 11 Q9OX23	Q9OX23 mus musculu
4	1587	33.4	635 4 Q96GX0	Q96GX0 homo sapien
5	941.5	19.8	457 4 Q9HBR1	Q9HBR1 homo sapien
6	897.5	18.9	296 11 Q9J129	Q9J129 ratus norv
7	876.5	18.5	963 4 Q9C0C4	Q9C0C4 homo sapien
8	860.5	18.1	893 4 Q9C0B8	Q9C0B8 homo sapien
9	697.5	14.7	762 4 Q9H3S1	Q9H3S1 homo sapien
10	688	14.5	687 4 Q9BXR8	Q9BXR8 homo sapien
11	686	14.5	1030 4 Q9H2E6	Q9H2E6 homo sapien
12	685.5	14.4	1049 4 Q9P2H9	Q9P2H9 homo sapien
13	685	14.4	284 11 Q54948	Q54948 mus musculu
14	682.5	14.4	1022 4 Q9P249	Q9P249 homo sapien
15	657	13.8	963 11 Q91Y36	Q91Y36 mus musculu
16	649	13.7	935 4 Q96JF8	Q96JF8 homo sapien

17	640	13.5	1005 11 Q9E071	Q9E071 mus musculu
18	639.5	13.0	287 4 Q9UDQ1	Q9UDQ1 homo sapien
19	618	13.5	517 4 Q9NRK9	Q9NRK9 homo sapien
20	610	12.9	724 5 Q9V707	Q9V707 drosophila
21	596.5	12.6	770 5 Q44253	Q44253 drosophila
22	596.5	12.6	770 5 Q9V3M4	Q9V3M4 drosophila
23	578	12.2	418 4 Q96J09	Q96J09 homo sapien
24	574.5	12.1	1202 4 Q9P283	Q9P283 homo sapien
25	561.5	11.8	328 13 Q9YHX3	Q9YHX3 brachydanio
26	541	11.4	367 4 Q9HAH9	Q9HAH9 homo sapien
27	520.5	11.0	616 5 Q9V7P8	Q9V7P8 drosophila
28	441.5	9.3	653 12 Q64906	Q64906 aleocephale
29	421.5	8.9	1083 5 Q9VTF0	Q9VTF0 drosophila
30	407.5	8.6	1081 5 Q9U631	Q9U631 drosophila
31	377	7.9	510 4 Q9NX92	Q9NX92 homo sapien
32	370.5	7.8	475 4 Q96FK5	Q96FK5 homo sapien
33	364.5	7.7	612 12 Q9J5F6	Q9J5F6 fowipox vir
34	332.5	7.0	658 5 Q95XP4	Q95XP4 caenorhabdi
35	323.5	6.8	658 5 Q9N3J8	Q9N3J8 caenorhabdi
36	322.5	6.8	424 4 Q9NS35	Q9NS35 homo sapien
37	321.5	6.8	676 5 Q9TYG4	Q9TYG4 caenorhabdi
38	305	6.4	699 4 Q96SM4	Q96SM4 homo sapien
39	225	4.7	1841 4 Q15031	Q15031 homo sapien
40	217.5	4.6	1894 11 P70206	P70206 mus musculu
41	209.5	4.4	328 4 Q9HA40	Q9HA40 homo sapien
42	207	4.4	264 5 Q95Q16	Q95Q16 caenorhabdi
43	190	4.0	1945 5 Q96681	Q96681 drosophila
44	190	4.0	1945 5 Q9V491	Q9V491 drosophila
45	189.5	4.0	1754 4 Q9U1W2	Q9U1W2 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9NS98	PRELIMINARY:	PRT:	782 AA.
AC	Q9NS98:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	SEMAPHORIN SEM2.			
GN	SEM2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Seki N., Hattori A., Hayashi A., Kozuma S., Muramatsu M., Miyajima N.,			
RT	Saito T.;			
RT	"Human semaphorin."			
RL	Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AB029486; BAB98132.1; -			
DR	InterPro; IPR003599; IG_1.			
DR	InterPro; IPR003006; IG_MHC.			
DR	InterPro; IPR003659; PSI.			
DR	Pfam; PF00047; Ig_1.			
DR	SMART; SM00409; IG_1.			
DR	SMART; SM00423; PSI_1.			
SO	SEQUENCE 782 AA; 86700 MW; 85CB424874DF6663 CRC64;			

Query Match	88.8%;	Score 4215;	DB 4;	Length 782;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 782;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	94 MAPSWAICWLGGLILHGSSGSPGSPVPRILSYRDLASNSAIRLFGQSLNQA	153		
DB	1 MAPSWAICWLGGLILHGSSGSPGSPVPRILSYRDLASNSAIRLFGQSLNQA	60		
OY	154 MWLEDRDLFLFGIDALSLRLDQAMPDPREVLPMPGQREECVKKGRDPLETCANFV	213		
DB	61 MWLEDRDLFLFGIDALSLRLDQAMPDPREVLPMPGQREECVKKGRDPLETCANFV	120		

```

QY 214 RVLQPHNRTHLLACGTGAFOPTCALITYGHRGEHVLHLEPGSVESGRGRCRHPSPRPAS 273
DB 121 RVLQPHNRTHLLACGTGAFOPTCALITYGHRGEHVLHLEPGSVESGRGRCRHPSPRPAS 180
QY 274 TFIIDELTYGLTADFLGDEAMIFRSGGPRPALRSDSDSLHDPFRVMAARIPENSDDN 333
DB 181 TFIIDELTYGLTADFLGDEAMIFRSGGPRPALRSDSDSLHDPFRVMAARIPENSDDN 240
QY 334 DKVYFEFSETYPSDPGSGNNHTVSRVGRVCVNDAGGQVLYNKKSTFLKARLYCSVPBG 393
DB 241 DKVYFEFSETYPSDPGSGNNHTVSRVGRVCVNDAGGQVLYNKKSTFLKARLYCSVPBG 300
QY 394 GAETHFDLEDFLLMPACKSLEYVALFSTVSAVFOGFAVCVYHMAIIMEVFENGFNR 453
DB 301 GAETHFDLEDFLLMPACKSLEYVALFSTVSAVFOGFAVCVYHMAIIMEVFENGFNR 360
QY 454 DGPQHQMGPTGKVPFPPRGVCPSPKMTAQPGRPFGSTKDYDDEVYLQFARAHPLMFWPVR 513
DB 361 DGPQHQMGPTGKVPFPPRGVCPSPKMTAQPGRPFGSTKDYDDEVYLQFARAHPLMFWPVR 420
QY 514 RHGRPVLYKTHLAQOLHOIYVDVRAEDGTVDYIFLGDSGSVLYKVALOAGGSAPEBEV 573
DB 421 RHGRPVLYKTHLAQOLHOIYVDVRAEDGTVDYIFLGDSGSVLYKVALOAGGSAPEBEV 480
QY 574 VLEELQYFKVPTITEMEISVYKRMVLYGSRGLVAQLRLHOCETTYGTACACCLARDPYC 633
DB 481 VLEELQYFKVPTITEMEISVYKRMVLYGSRGLVAQLRLHOCETTYGTACACCLARDPYC 540
QY 634 AMDASCTHYRPSLGKRRFRRODIRHGNPALQCLGQSOEEBAVGLVATWVYGTENSTF 693
DB 541 AMDASCTHYRPSLGKRRFRRODIRHGNPALQCLGQSOEEBAVGLVATWVYGTENSTF 600
QY 694 LECIPKSPQAAVRLRLORPGDEGPDQYKTDERVLHTEGILFRSLRSDACTYCTTLEH 753
DB 601 LECIPKSPQAAVRLRLORPGDEGPDQYKTDERVLHTEGILFRSLRSDACTYCTTLEH 660
QY 754 GFSQTVVRLALVIVASQDLNLFPEPEKPEEPARGGLASTPPKAWYKDILQILGFA 813
DB 661 GFSQTVVRLALVIVASQDLNLFPEPEKPEEPARGGLASTPPKAWYKDILQILGFA 720
QY 814 RVDEYICRWRCRGTTGSGCFRSGRSGKQARGKSMAGLELGKKMSRYHAHNTPREVE 873
DB 721 RVDEYICRWRCRGTTGSGCFRSGRSGKQARGKSMAGLELGKKMSRYHAHNTPREVE 780
QY 874 AT 875
DB 781 AT 782

```

RESULT 2

```

Q9H703 PRELIMINARY; PRT; 725 AA.
AC 09H703;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE FLJ00014 PROTEIN (FRAGMENT).
GN FLJ00014.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RL spleen."
DR EMBL; AK024425; BAB15715.1;
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig_like.

```

```

DR InterPro; IPR003606; Ig_MHC.
DR InterPro; IPR003659; PSI.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00409; Ig_1.
DR SMART; SM00410; Ig_like_1.
DR SMART; SM00423; PSI_1.
FT NON_TER
SQ SEQUENCE 725 AA; 80801 MW; 8EC8B88FC580D743 CRC64;

```

```

Query Match 82.4%; Score 3913; DB 4; Length 725;
Best Local Similarity 100.0%; Pred. No. 5,5e-317;
Matches 725; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 151 LQAMLYDERDLFLFGIDLALSLRLDQAMPDREVLMPDPQGRGECVRGRDPLETCA 210
DB 1 LQAMLYDERDLFLFGIDLALSLRLDQAMPDREVLMPDPQGRGECVRGRDPLETCA 60
QY 211 NFVRVLOPHNRTHLLACGTGAFOPTCALITYGHRGEHVLHLEPGSVESGRGRCRHPSPRP 270
DB 61 NFVRVLOPHNRTHLLACGTGAFOPTCALITYGHRGEHVLHLEPGSVESGRGRCRHPSPRP 120
QY 271 FASTFIDELTYGLTADFLGDEAMIFRSGGPRPALRSDSDSLHDPFRVMAARIPENS 330
DB 121 FASTFIDELTYGLTADFLGDEAMIFRSGGPRPALRSDSDSLHDPFRVMAARIPENS 180
QY 331 QDNKVVYFFSETYPSDPGSGNNHTVSRVGRVCVNDAGGQVLYNKKSTFLKARLYCSVP 390
DB 181 QDNKVVYFFSETYPSDPGSGNNHTVSRVGRVCVNDAGGQVLYNKKSTFLKARLYCSVP 240
QY 391 GPGAEETHFDLEDFLLMPACKSLEYVALFSTVSAVFOGFAVCVYHMAIIMEVFENGF 450
DB 241 GPGAEETHFDLEDFLLMPACKSLEYVALFSTVSAVFOGFAVCVYHMAIIMEVFENGF 300
QY 451 AHRGPOHQMGPTGKVPFPPRGVCPSPKMTAQPGRPFGSTKDYDDEVYLQFARAHPLMFWP 510
DB 301 AHRGPOHQMGPTGKVPFPPRGVCPSPKMTAQPGRPFGSTKDYDDEVYLQFARAHPLMFWP 360
QY 511 VRPRHGRPVLYKTHLAQOLHOIYVDVRAEDGTVDYIFLGDSGSVLYKVALOAGGSAPE 570
DB 361 VRPRHGRPVLYKTHLAQOLHOIYVDVRAEDGTVDYIFLGDSGSVLYKVALOAGGSAPE 420
QY 571 EEVVLQYFKVPTITEMEISVYKRMVLYGSRGLVAQLRLHOCETTYGTACACCLARD 630
DB 421 EEVVLQYFKVPTITEMEISVYKRMVLYGSRGLVAQLRLHOCETTYGTACACCLARD 480
QY 631 PYCAMDASCTHYRPSLGKRRFRRODIRHGNPALQCLGQSOEEBAVGLVATWVYGTEN 690
DB 481 PYCAMDASCTHYRPSLGKRRFRRODIRHGNPALQCLGQSOEEBAVGLVATWVYGTEN 540
QY 691 STFLECLPKSPQAAVRLRLORPGDEGPDQYKTDERVLHTEGILFRSLRSDACTYCTT 750
DB 541 STFLECLPKSPQAAVRLRLORPGDEGPDQYKTDERVLHTEGILFRSLRSDACTYCTT 600
QY 751 LEHGFQTVVRLALVIVASQDLNLFPEPEKPEEPARGGLASTPPKAWYKDILQILGFA 810
DB 601 LEHGFQTVVRLALVIVASQDLNLFPEPEKPEEPARGGLASTPPKAWYKDILQILGFA 660
QY 811 NLRPVDEYICRWRCRGTTGSGCFRSGRSGKQARGKSMAGLELGKKMSRYHAHNTPR 870
DB 661 NLRPVDEYICRWRCRGTTGSGCFRSGRSGKQARGKSMAGLELGKKMSRYHAHNTPR 720
QY 871 EVENT 875
DB 721 EVENT 725

```

RESULT 3

```

Q9QX23 PRELIMINARY; PRT; 775 AA.
AC 09QX23;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)

```

DE SEMAPHORIN M-SEMAK.
 GN SEMA3E.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57/BLACK 6.
 RA Miyazaki N., Furuyama T., Inagaki S.;
 RT "A novel semaphorin, M-Semak which inhibits neural outgrowth from
 RT sensory neurons."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF034744; AD01996.1; -
 DR MCD; MGI:1340034; Sema3e.
 DR Interpro: IPR003599; Iq.
 DR Interpro: IPR003006; Iq_MHC.
 DR Interpro: IPR003659; PSI.
 DR Interpro: IPR001627; Sema.
 DR Pfam: PF00047; Iq.1.
 DR Pfam: PF01403; Sema.1.
 DR SMART: SM00409; Iq.1.
 DR SMART: SM00423; PSI.1.
 DR SMART: SM00423; PSI.1.
 SQ SEQUENCE 775 AA; 89543 MW; 221E766F404098D4 CRC64;

Query Match 43.2%; Score 2052.5; DB 11; Length 775;
 Best Local Similarity 50.7%; Pred. No. 4,6e-162;
 Matches 399; Conservative 127; Mismatches 234; Indels 27; Gaps 10;

OY 94 MAPSAMICWLLGILLHGGSSGSPGSPVRLRLSYRDLISANSALFPGSLNIOA 153
 DB 1 MAPAGHITLLMLGHLLEMTLPGHSANPSYRLRLSHKLELNTSTJFOSLGLDHT 60
 OY 154 MYLDEYRDLRLGLDALYSURLDOAMPDPREVLMPPQPGOREECVRGRDPLTECANFV 213
 DB 61 MLDEYORLRLVGGRDLYSLNLEKVSQYREITWPSYAVKVECTIMKGD-ANCAANI 119
 OY 214 RVLORPHNTHLLACGTGAFOPTCALITVGHGHEVL-HLEPGSVESGRCRCHPERSRPA 272
 DB 120 RVLHNYNTHLLTCAFGAFCATVGHGHEEPLFHLESIRSGRCRCPEDNSFV 179
 OY 273 STFDIGELYTLADFLGREAMIFRSGRPRLRSD-SDQSLHDPKRYMAARITENSQ 331
 DB 180 STLVNELLFAGLYSDYWRGDSAFIRSMGKLGIRTEHDERLKEPKFVSGIMRIDNER 239
 OY 332 DNDKYEFSEFVPSPDGGSNHTVSRVGRVCVNDAGGGRVLYNKMSTFLKARLYCSVP 391
 DB 240 DNDKYTEFSEFVPSPDGGSNHTVSRVGRVCVNDAGGGRVLYNKMSTFLKARLYCSVP 298
 OY 392 PGGAETHFDQLEDELFLMPKAGKSLEYVALFSTVSATVPGFAGVAVYHMDIWEVNGP 451
 DB 299 MNGIDTYDELEDVFLPTROPKNPVIGLFNTSIFRGHNAVYVHMSIREAFNGPYA 358
 OY 452 HRDQPOHOMGRGKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMFV 511
 DB 359 HKEGGEYHMSLGEKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMFV 415
 OY 512 RRRGRPVLYKTHLAQOLHQLVVDVREAEADGTVDVFLTSDGSLVKYALDAGSABE 571
 DB 416 RYVHKKPIIVKTDGKNYRLAVDRVDAEDGQYDLFTDIDGYLAKVITITINOETEMNE 475
 OY 572 EYVLEELQVFKVPTITEMEISVKRQMLVGSRLGVALRLHQCETTYGACACCLARP 631
 DB 476 EYVLEELQVFKVPTITEMEISVKRQMLVGSRLGVALRLHQCETTYGACACCLARP 535
 OY 632 YCAMDGASCTHNR--SLGKRERRRDIHNGRPAIQCIGOSOEAEVGLVATVNYGTE 689
 DB 536 YCAMDGASCTHNR--SLGKRERRRDIHNGRPAIQCIGOSOEAEVGLVATVNYGTE 595
 OY 690 NSTFLECLPKSPQAAVRLQLRPGDEGPDQVYTERDEVLTENGILFRRLSRPDAGTYTCT 749
 DB 596 NSTFLECLPKSPQAAVRLQLRPGDEGPDQVYTERDEVLTENGILFRRLSRPDAGTYTCT 655

OY 750 TLEHFSQTVRLALVIVASQIDNLPPEPKPE-----PPARGGLASTPPKAMTKDI 803
 DB 656 TVEHNFVHTVAKITLLEVVEHKEVMEHKKHDEERHNMPCPPLSGMSQGTKP--WYKEF 713
 OY 804 LQIGFANPLPVDEYCEVWCRGTTECGCFRSGRCGQAGKSMAGLELKKKSRVHA 863
 DB 714 LQIGFANPLPVDEYCEVWCRGTTECGCFRSGRCGQAGKSMAGLELKKKSRVHA 763
 OY 864 EHNRRPR 870
 DB 764 EHNRRPR 770

RESULT 4
 ID 096GXO PRELIMINARY; PRT; 635 AA.
 AC 096GXO;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:18122)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRIN, AND ANAPLASTIC OLIGODENDROGLIOMA WITH 1p/19q LOSS;
 RA Strausberg R.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC009113; AA09113.1; -
 DR EMBL: BC009113; AA09113.1; -
 SQ SEQUENCE 635 AA; 70694 MW; D994099B476B9210 CRC64;

Query Match 33.4%; Score 1587; DB 4; Length 635;
 Best Local Similarity 49.9%; Pred. No. 2e-123;
 Matches 321; Conservative 90; Mismatches 196; Indels 36; Gaps 12;

OY 211 NFRVLOPHNTHLLACGTGAFOPTCALITVGHGHEVL-HLEPGSVESGRCRCHPERSR 269
 DB 2 NFRVLOPHNTHLLACGTGAFOPTCALITVGHGHEVL-HLEPGSVESGRCRCHPERSR 61
 OY 270 PFASTFIDGELYTLADFLGREAMIFRSGRPRLRSD-SDQSLHDPKRYMAARITENSQ 328
 DB 62 PFASTFIDGELYTLADFLGREAMIFRSGRPRLRSD-SDQSLHDPKRYMAARITENSQ 121
 OY 329 SPQDNDKYEFSEFVPSPDGGSNHTVSRVGRVCVNDAGGGRVLYNKMSTFLKARLYCS 388
 DB 122 SPQDNDKYEFSEFVPSPDGGSNHTVSRVGRVCVNDAGGGRVLYNKMSTFLKARLYCS 181
 OY 389 VPFGAETHFDQLEDELFLMPKAGKSLEYVALFSTVSATVPGFAGVAVYHMDIWEVNG 448
 DB 182 VPFGAETHFDQLEDELFLMPKAGKSLEYVALFSTVSATVPGFAGVAVYHMDIWEVNG 239
 OY 449 PRAHNRGPOHOMGRGKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMF 508
 DB 240 PRAHNRGPOHOMGRGKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMF 295
 OY 509 WPRHNRGPOHOMGRGKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMF 568
 DB 296 WPRHNRGPOHOMGRGKVPYRPGVCPSCMTAOPRGPGSKYDPRDEVLOARAHPLMF 355
 OY 569 EPEVLEELQVFKVPTITEMEISVKRQMLVGSRLGVALRLHQCETTYGACACCLARP 628
 DB 356 EPEVLEELQVFKVPTITEMEISVKRQMLVGSRLGVALRLHQCETTYGACACCLARP 415
 OY 629 RDPYCAMDGASCTHNR--SLGKRERRRDIHNGRPAIQCIGOSOEAEVGLVATVNYGTE 688
 DB 416 RDPYCAMDGASCTHNR--SLGKRERRRDIHNGRPAIQCIGOSOEAEVGLVATVNYGTE 470
 OY 689 HNSTFLECLPKSPQAAVRLQLRPGDEGPDQVYTERDEVLTENGILFRRLSRPDAGTYTCT 748
 DB 471 HNSTFLECLPKSPQAAVRLQLRPGDEGPDQVYTERDEVLTENGILFRRLSRPDAGTYTCT 530

[illegible]

DR SMART: SM00409: IG: 1.
 DR SMART: SM00410: IG: like: 1.
 DR SMART: SM00423: PSI: 1.
 FT NON_TER 1
 SO SEQUENCE 963 AA; 106735 MW; 3FA37DBA3483ECE CRC64;

Query Match 18.5%; Score 876.5; DB 4; Length 963;
 Best Local Similarity 32.2%; Pred. No. 4.9e-64;
 Matches 256; Conservative 108; Mismatches 295; Indels 135; Gaps 32;

40 SLLSSAP-----LPADWVEPLPYKMWPGSGSRANYNRRPAGGSGAGRR--QRC 88
 76 SLWSSGRKMLKRYPSFLPAA-WICLLP--GW-----ERLGRPMGCGORLFCXC 123
 89 POPSAPSAW-AICWLLGLLHGGSSGP-----SPGPSV-RRLSLYRLLSANBSA 140
 124 PLRP-IRGFGWLLVAM-----GAGSGARLRAVEPGSCSAAMLPAELATVYRF 175
 141 IFGLPGSLNLQAMVLDKXRDRLFLGLDALYSRLDQ-----AMPDPREVLMPDQ 193
 176 SQGIDQFLTLT---LLEPTGLLYGAREALFAFSMEALELGALISWEAPVE----- 224
 194 QRECVARKGDPTECANRYVLOPHNRTHLACGTGAPOPTCALITVGHGHEVHLER 253
 225 KTECLOKGNKNOTECFNFIFLOPYNASHLVYCGTYAQPKCTVNM-----LRETLER 279
 254 GVSSEGRCPHPSRPFATFIDGELYGLTADFLGREAMIFRSGPPRLSRDSDQSL 313
 280 GEEEDGKCPYPAKAGHALLVDGELYSATLNNFLEGTETILRNNGPHMSKTEYLAW 339
 314 LHPREFMARIPENDQ---DNDKYFFPSETVPSPDGSNHYTVSRGRCVNDAGQ 370
 340 LNEHFVGSAYVESVSGTFGDDKVFEPFERAVESDCYAEQV-VARVAVCKGMDGA 398
 371 RVLNKKSTLKLARLVSVGPGCAETHPDQEDVFLMKKAKSLEYVLFSTVSAVPG 430
 399 RTLOKRTTLKRLKACSA---NMOLYFNOLQAMHTLQDTSMTNTTFEGVFAQWQDA 455
 431 GFACVVMHMDIMEVFNPGRAHRDGPQHGPGYKVPFRPGVCPSKMTAOCGRPFST 490
 456 LSAICEVQLEIQLVFGFPIKEHEEAKMDRYTDPSPRPGSCINNMHRRNG--YTSS 513
 491 KDYDEVLOFARAHLMFMVPRHGRPLVK-----THLAQOLHOIVDREAEDG-TY 544
 514 LELPDNLINFKRHLMEQVGRMSRPLVKKGTNFTHL-----VADRVGLDQATY 566
 545 DVTGLDGSGLKVLIALQAGSAPREEVILEEQVKVPTITMEISYKROMLYVGR 604
 567 TVLEIGTGDGWLKAVSL-----GPMVHLELEQLD-OEPMSLVSOSKLLFAGSR 619
 605 LGVALRLHOCETGTACAECCCLARDPYCAMDASCTHYRPSLGRFRRODIRHGNPAL 664
 620 SOLVQLPAPADOMK-RSADQCVLARDPYCAM-----SVNTSCVAVGHSGLLI 668
 665 QCLGQSOEEBAVGLVATMYGTENST-----FLECLPKSPAANRML---QRRPD 714
 669 QHVMSTDSGICNIRGSKRYRTPKNTIVVAGTDLVPCHLSSMLAIAKMTFGGRDPAE 728
 715 EGPQGVKTEDEVILHERLIFRRLSRPDAGTYCTTLEHGSQTVVRLA---LVVIVA- 769
 729 QPGSFLYDARL-----QALVYMAAPRHAAGYHCFSEBQ-----ARLAAGYLIVAVAG 778
 770 -----SOLDNL 775
 779 PSVLEARAPLENL 792

RESULT 8
 Q9C0B8 PRELIMINARY: PRT: 893 AA.
 AC Q9C0B8:
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE K1A1745 PROTEIN (FRAGMENT).
 GN K1A1745.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 7:347-355(2000).
 DR EMBL: AB051532; BAB21836.1; -;
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR003659; PSI.
 DR Pfam: PF01437; Plexin_repeat; 1.
 DR SMART: SM00423; PSI: 1.
 FT NON_TER 1
 SO SEQUENCE 893 AA; 98361 MW; D6C6C48DEE524F14 CRC64;

Query Match 18.1%; Score 860.5; DB 4; Length 893;
 Best Local Similarity 31.4%; Pred. No. 9.5e-63;
 Matches 253; Conservative 109; Mismatches 306; Indels 139; Gaps 31;

37 QPSLLSSAPLPADWVEPLPYKMWPGSGSRANYNRRPAGGSGAGRRCPQPPSAP 96
 3 QGFLDPVSHLPPRS-----CG-----GPGRGDGA--DRGAELPVPSP 39
 97 SAMWICWLLGLLHGGSSGSPGPS---VPLR-----LSYRLLSANBSAI----- 141
 40 -----APPEPERDVTVAALRLKRAMGRSLWALPWGALPRLPPL 82
 142 -----FLGPGSLN-----QAMVLDKXRDRLFLGLDALYSRLDQ-----RD--RLPLG 167
 83 LLLILLILLQPPPTWALSPLSLPSGERPFLKFAEHSNTYALLLSRGRLLYGA 142
 168 LDALYSRLDQAM---PDPREVLMPDQPGQRECVARKGDPTECANRYVLOPHNRTHL 224
 143 REALFALSSNLSPLPGVEYELLMGDAERKQOCFSFKGKDPQDQNYIKILLPLSGHL 202
 225 LACGTAPOPTCALITVGHGHEVHLERGSV--ESGRCPHPSRPFASFIDGELYT 282
 203 FTCGTAFSPMCTYINM--ENFTLADKERGNVLEDEGKRCFPDPVSTALVDGELYT 260
 283 GLPADFLGREAMIFRSGGPPALRSDDQSLHDPREFVMAARIPENDQ---DNDKYFF 339
 261 GYVSSFOGNDPALRSQSLRPT-KTESSLNWLQDPAFVASAVIPESLGLQDDDKIYFF 319
 340 FSETVSPDGSNHYTVSRGRCVNDAGGORYLVNKKSTFLKARLYCVGPGCAETHF 399
 320 FSETGOEFFEFTNT-VSRIARICKDEGEERYLQORWTSFLKAOLLCRDPDG--PPE 375
 400 DQLEDEVEL--WPKAKSLEYVALFSTV--SAVQSGFCVVMHMDIMEVFNPGRAHRD 455
 376 NVLDQVFTLSPQDMKMDLFTGVTSQMHRTGTSASAVCFMKRYQVFGSLYKEVNR 435
 456 POHMGPGYKGVPPRPGVCPSTMTAOPGRPGSTKDYDEVLOFARAHLMFMVPRH 515
 436 ETQQWYVTHPVPPLPRGACITNSARE--RKINSSQLDPRVNLFLKHDGQVNR-- 490
 516 GRPYLAKTHLAQOLHOIVDREAEDGTYVIFLGDSSGLKVLIALQAGSAPREEV 575
 491 SRMLLOPARYO--RVAHVRVGLHHTYDVFLGTDGGLKAV-----SYGPRVHTI 542
 576 BELQVEKVPPTITEMEISYKROMLYVGRSLGVALRLHOCETGTACAECCCLARDPYCAM 635
 543 BELQIESSGQPVONLLDITRGLLYAASHGQVQVAMANSLSL-RSCGOLLARDPYCAM 601
 636 DGASCTH---YRSLGKRFRFRRODIRHGNPALQCLGQSOEEBA---VGLVATMYGTENH 689

Db 602 SSSCKHVSIXPOLATRPW-IQDIEGASAKDLCASSVSPSPVTEGKPCQEQVQFQPN 660
 QY NSTFLECPKSPQAAVRMLQRPDEGPOVKTDERVLHTEGGLFRRLSRDPACTYTC 749
 Db 661 TWTNLTACCLSLATRRMLD--RNG--APYNASASCHVLTGDL--VGTOOLEFQCM 713
 QY 750 TLEHGSQTVVRLALVIT--VASOLD 773
 Db 714 SLEEGFOOLVASYPEVEDEVDADOTD 740

RESULT 9

Q9H3S1 PRELIMINARY; PRT; 762 AA.
 ID Q9H3S1
 AC Q9H3S1
 DT 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 DE SEMB.
 GN SEMB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Seki N., Hattori A., Hayashi A., Kozuma S., Muramatsu M., Miyajima N.,
 RA Saito T.,
 RT "Human semaphorin B."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB029394; BAR20087.1;
 DR InterPro: IPR000005; HTHAFC.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR003659; PSI.
 DR Pfam: PF01437; Plexin_repeat; 1.
 DR PRINTS: PR00142; RECA.
 DR SMART: SM00423; PSI; 1.
 DR PROSITE: PS00041; HTH_ARC_FAMILY_1; UNKNOWN; 1.
 DR SEQUENCE 762 AA; 83195 MW; 6A738E4EBEB834D4 CRC64;

Query Match 14.7%; Score 697.5; DB 4; Length 762;
 Best Local Similarity 28.4%; Pred. No. 2.8e-49;
 Matches 236; Conservative 125; Mismatches 326; Indels 145; Gaps 35;

QY 92 PSAPASMAIWMILGILLH-----GSSGSPSPSPVRLSLSYRDLSSANRSATF 142
 Db 4 PALGLDPWMLGLFLFOLLPTTATAGGGGCGP-----MPVRYAGD---ERRALSF 55
 QY 143 LGPGSLNQAMYLDEYRDLFLGGLDALYSLY-DOAMPDPREVLPWPGQORECVR 200
 Db 56 FHOKGLQDEDTLLSGDGNLTVGAREALIALDIOPGVRLKMLPWPASDRKSKSECAF 115
 QY 201 KGRDPLTECANFVRLQPHNRTHLACGAFQPTCALIVGHEVHLLEGSVSSGR 260
 Db 116 KKSNETQCFNFRVLVSVNTHLYLCGTFAPFACFTFEL--QDSYLLTISEDKWEKG 173
 QY 261 GCRPHSPSPASTFDGLYLTGADFLGREAMIFRSGGPRALSDSD-QSLHDPFR 319
 Db 174 GQSPEDPAKHHTAVLDGMLYSGTNNFLGSEPIILMTLGSQVLTNDNLRMLHNDASF 233
 QY 320 VAAARIPESDQNDNVYFEFSTVSPSPGSGNHVSVRGVCVNDAGQARVLVWKMT 379
 Db 234 V--AATP-----STGVYVFFETETASEPD-FPERLHSTRVARKCKNDVGKLLQKKWT 285
 QY 380 FLKARLVCSVPPGGAETH-----FDQLEDFVFLMPKAGKSLLEVYALFVSVA 428
 Db 286 FLKADLT-SAPRGSCSPSSATRSCTPILQLPTSTQSSPESG-----VGTRSS- 337
 QY 429 FCGFVAVYHMDIMEVFNPRFAHNDPOHONGVPGKVPFRPGVCSPKMTAQDPGPGA 488
 Db 338 ---AVCAFLSLDIEVRFKKEKELNKTSTRTVYRGPEINRPGSC----- 380

QY 489 STRKYDPDEVLOFARAHPLMFVPRPHGRVLYKTHLAQOLHIVDYRVEADG--TYDYI 547
 Db 381 SVGPSDDKALTFMKDHFLEDEGV---VGPRLVKSQV--EYRLAVETAQGLDGHSLWM 435
 QY 548 FLGTDSGLVKYIALQAGSAPDEEVLELOVFNVPRTITMEISVKRQMLYVSRGCV 607
 Db 436 YIGTTTGSILKRAVY-----SGDSSAHVLEIOLEPDEVRNLQAPQGAFAVFGSGV 490
 QY 608 AOLRIHOCETVGTACAECCLPARDPYCAMDGAS---CTHYRPSLGRFRFRODIRGNAL 664
 Db 491 WRVPRANCSTV-ESCVDCVLARDPCAMPDESRTCCLSAPRLNS---WKQMEGNENW 546
 QY 665 QCLGSGQEEANGLVAATM-----YGTGHNSTFLECLPKSPQAAVRMLQRP 713
 Db 547 AC-----ASGPRSRSLRQSPRIIKEYLAVNSLILEPCPHLSALASYW----- 592
 QY 714 DEGPQVKTDERVLTHTEGGLFRRLSRFDAGYTCCTLEHGSQTVVRLALVITASOLD 773
 Db 593 SHGPAAYPEASTVY--NGSILLIYQDVGGLYOCWATENGSTYVLS---YWDSD-D 645
 QY 774 NLFPPEPKPEEPARGGLASTPPKAMYDIIQLTGFMILPRVDEXCEVWCNGT----- 828
 Db 646 QTLADPE-----LAGIPREHVKVLTRVSGGALAAQGSY---WHFVTVVLF 692
 QY 829 --ECSCGF-----RSRSRQKARGK--SWAGLELGKKK--SRVHAENHRTPRE 871
 Db 693 ALVLSGALLIIVASPLRALRARGKVGCGETLRPEKAPLSR--EQLHOSPRE 742

RESULT 10

Q9BX8 PRELIMINARY; PRT; 687 AA.
 ID Q9BX8
 AC Q9BX8
 DT 01-JUN-2001 (Tremblrel, 17, Created)
 DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 DE SEMAPHORIN 6B ISOFORM 2.
 GN SEMA6B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21248680; PubMed=11350127;
 RA Correa R.G., Sasahara R.M., Bengtson M.H., Katayama M.L.H.,
 RA Salim A.C.M., Brentani M.M., Sogayar M.C., de Souza S.J.,
 RA Simpson A.J.G.,
 RT "Human Semaphorin 6B (HSA) SEMA6B", A Novel Human Class 6 Semaphorin
 RT Gene: Alternative Splicing and All-Trans-Retinoic Acid-Dependent
 RT Downregulation in Glioblastoma Cell Lines.";
 RL Genomics 73:343-348(2001).
 DR EMBL: AF293363; AK16831.1;
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR000737; Squash.
 DR SMART: SM00423; PSI; 1.
 DR SMART: SM00286; PTL; 1.
 DR SEQUENCE 687 AA; 74967 MW; CBG68CF5D69738D7 CRC64;

Query Match 14.5%; Score 688; DB 4; Length 687;
 Best Local Similarity 32.0%; Pred. No. 1.5e-48;
 Matches 203; Conservative 91; Mismatches 262; Indels 78; Gaps 24;

QY 90 QPFSMAPSMAIWMILGILLHGGSSGSPSPVRLSLSYRDLSSANRSATFGLG----- 144
 Db 2 QTPRAPPRLALL--LTLGGAHGLFP-EPPPLSVAPRDYL--NHYPVFGSGPGR 55
 QY 145 ---PQSLNQAMYLDEYRDLFLGGLDALYSLRLQAMPD-----PREVLWPPQPGQRE 197
 Db 56 LTPAEGADDLTIQVRLVNRVTLFIDGRDNLRYVELEPPITSTELRYOKLTWNSPESDIN 115
 QY 198 CYRKGADPLTECANFVRLQPHNRTHLACGGAFOPTALITVGRGHVHLLEP-GSV 256

```

Db 116 CRMKGKE-GEGRNFVKYLLRDESTLFCGSMNFPCANYSID-----TLQPVGDN 167
QY 257 EGRGRCHREPRRPASTFICGLTGLADFLGREAMIFRSGRPPALRS-DSQSLIH 315
Db 168 ISGMARCPDRPHANVALESDMLTATVDTFLADAVIYSLGRPTLRVYKHDKMKFK 227
QY 316 DREFVMAARIPENSDDNDKYFFPSETVPSPDGGSNHVTYSRVGRVCVNDAGO-RVLY 374
Db 228 ERYFYAAV-----EMSHYVEFFRE-IAMFNYLEKVVSRVAVCKNDVGGSPVLE 279
QY 375 NKWSTFLKRLVCSVPGGGAETHDQLEDFLLMPKAGKSLFVYALFSTYSAVPOGPAV 434
Db 280 KQWSTFLKRLVCSV--PQDSHFYFENVLDVAVGVSLGSRV-VLAIVSTPNSITPGSAV 336
QY 435 CVYHMAIDIEVFNQFPAHNDGPOHMGPR-GGKVPFRPGVCPSKMTAOPGRPFSTKDY 493
Db 337 CAFDLVQVAVVFGRRRECKSPESITVPRBEDQVPRPRGCC-----AARGMAYNASAL 391
QY 494 PDEVLOFARAHPLMPVPRRHRGRLVYKTHLAQQLHQIVDVRVLEADGTVDYIFLGTD 553
Db 392 PDDLNFVKTHPLMDAVALVSLGHAPWILRLTHOLTRVADVAGAPMGQVYVFLGSEA 451
QY 554 GSVLKYV-----ALQAGSAPREVEVLEQVFK-----VPTPTEMEIS 593
Db 452 GTVLKFLVRPNAVSTSGTSG--LSVPLEEFETVPRDRCERPGGETGQRLDLSLEID 509
QY 594 VKRQMLYVGRSLVAVQLRLHQIETGTAECCLARDPYCAW--DGASCTHYRPSLGKR 651
Db 510 AASGGLAARPCVVRVAVRARCQOYSGCMKNCIGSDPYCGMAPDG-SCIFLSP--GTRA 566
QY 652 FRRODIRHGNPA--LQCLG---QSOEEAVGLYA 680
Db 567 AFEQDVAGASTSGIGDCTGLLRASLSEDRAGLYS 600

```

RESULT 11

```

ID 09H2E6 PRELIMINARY; PRT; 1030 AA.
AC 09H2E6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SEMAPHORIN SEMA6A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564339; PubMed=10993894;
RA Klostermann A., Lutz B., Gerlier F., Behl C.;
RT "The orthologous human and murine semaphorin 6A-1 proteins (SEMA6A-
RT 1/SEMA6A-1) bind to the Enabled/vasodilator-stimulated Phosphoprotein-
RT like Protein (EVL) via a novel carboxyl-terminal zyxin-like domain."
RT J. Biol. Chem. 275:39647-39653(2000).
RL EMBL: AF279656; AAC29378.1; -.
DR InterPro: IPR003659; PSI.
DR SMART: SM00423; PSI.
SQ SEQUENCE 1030 AA; 114368 MW; A57B79C10AEC4B34 CRC64;

Query Match 14.5%; Score 686; DB 4; Length 1030;
Best Local Similarity 30.3%; Pred. No. 3,9e-48;
Matches 199; Conservative 93; Mismatches 278; Indels 86; Gaps 24;
QY 104 LLGGLLHGSSGSPGSPVRLRLSYRDLISANRSAIFLGPGSLN-----LQAMYLD 157
Db 7 LLFTLLHFAAGG--PPEDEP--ISISHGNV--TKQYPVVGHKPGRNITQRRHLDIOMIM 62
QY 158 EYRDLRFLGGLDALYSRLDQAMPD-----PREVLMPPGRGRECVRRKRDPLTECANV 213
Db 63 IMMGTLYIARNDHIYVDITSTHEIYCSKLTWKSROADVDTCKMKGKHK--DECHNEFI 121
QY 214 RVLDPRHNRHTLLACGTCGAPRPTCALITVGHGRGHVHLDEP--GSVEGRGRCHREPRRPA 272

```

```

Db 122 KVLKKMDNDLFLVCGTNAFNPSC-----RNYKMDTLEBFGEFEGMARCPDANKANV 174
QY 273 STFIDGLYGLRADFLGGRAMIFRSGRPPALRS-DSQSLIHDRFVMAARIPENSQ 331
Db 175 ALFADGKLSATVYDFLDAIVAYRSLGESPTLRKYKHDKMLKEFYQAV-----D 227
QY 332 DNDVVFYFSPVSPDGGSNHVTYSRVGRVCVNDAGG--ORVLNWKSTFLKRLVCSVP 390
Db 228 YGDIYFFERE--IAYENITMGKVPRVAVQCKNDMGSGRLKQWTSFLKARLNCVAV 286
QY 391 GPGGAETHF-----DQLEDFLLMPKAGKSLFVYALFSTYSAVPOGPAVYHMAIDIEV 445
Db 287 G-----DSHFYENIILQAVTDVTRI--NGRDV-VLATSTPYNSTPGSAVYMDLIASV 338
QY 446 FNGRPAHNDGPOHMGPR-GGKVPFRPGVCPSKMTAOPGRPFSTKDYPRDEVLOFARAH 504
Db 339 FTGRRECKSPDSTVTPVPRDERVPRKPRGCCASSSILE-----RVATSNRPDDTLNFIKTH 395
QY 505 PLMPVPRRHRGRLVYKTHLAQQLHQIVDVRVLEADGTVDYIFLGTDSSVLYKVALQA 564
Db 396 PLMDAVALVSLGHAPWILRLTHOLTRVADVAGAPYQNHVYVFLGSEKGLITFLRIG 455
QY 565 GGSAPREVEVLEQVFK-----VPTPTEMEISVKRQMLYVGRSLVAVQLRLHQ 615
Db 456 NSGFLNDSIFLEEMSVYSEKCSYDGVEDKRIHQMDRASSLSLYAFSTVIVPLIGRC 515
QY 616 EYTGTAACAECCLARDPYCAW--DGASCTHYRPSLGKRFRRODIRHGNPAIQLQSOEE 673
Db 516 ERHGCKKTKIARSPRYCGMKEGAGSHLSPN--SRLEFDIEDIRGN-----T 562
QY 674 EAVGLVATVYGTETHNSTFLFLECLKSPQAA-----VRM--LLQRPDESP 717
Db 563 DGIQVCHNSFVALNGHSSSLPSTTSDSTAQEGYESRGKMLDMKHLIDSPDTPR 618

```

RESULT 12

```

ID 09P2H9 PRELIMINARY; PRT; 1049 AA.
AC 09P2H9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KIAA1368 PROTEIN (FRAGMENT).
DE KIAA1368.
GN KIAA1368.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRAIN;
RC MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 7:65-73(2000).
DR EMBL: AB037789; BAA92606.1; -.
DR InterPro: IPR003659; PSI.
DR SMART: SM00423; PSI.
FT NON_TER
SQ SEQUENCE 1049 AA; 116511 MW; 7781D20ACC7A8AEA CRC64;

Query Match 14.4%; Score 685.5; DB 4; Length 1049;
Best Local Similarity 30.3%; Pred. No. 4,4e-48;
Matches 202; Conservative 94; Mismatches 280; Indels 91; Gaps 26;
QY 104 LLGGLLHGSSGSPGSPVRLRLSYRDLISANRSAIFLGPGSLN-----LQAMYLD 157
Db 9 LLFTLLHFAAGG--PPEDEP--ISISHGNV--TKQYPVVGHKPGRNITQRRHLDIOMIM 64
QY 158 EYRDLRFLGGLDALYSRLDQAMPD-----PREVLMPPGRGRECVRRKRDPLTECANV 213

```

```

Db 65 IMNCTLYIAADHITYVDIDTSHTEIYCSKLTGWSQADVIDORMGKHK-DECHNFI 123
QY 214 RVLDPHNTHLLAGTGAFQPTCALITYGHGHEVHLIEP-GSVESGRCRPHSPFA 272
Db 124 KVLKKKDDALLFCVGTANAFNSC-----RNYKMDTLEPDESGMARCPYDAKIHANV 176
QY 273 STEIDELTYGLADFLGREAMIFRSGGPPALRS-DSDSGLHDPFRVMAARIPENDQ 331
Db 177 ALFADGKLYSATVDTFLAIDAVIYRSLGESPTLTKVHKSWMLEPFTVQAV-----D 229
QY 332 DNKVFEEFSTVSPGSGNHTVSRGVYCVADAGC-QRVLVNKMSTFLKARLVCSVP 390
Db 230 YGDIYTFEERE-LAVEYNMGKVFEPVAVQCKNDMGSGORLEKOWTSPFLKARLVCSVP 288
QY 391 GPGAEATHF-----DOLEDFLLMPKAGKSLEYALFSTVASVFGAVCYVHMADIWEV 445
Db 289 G-----DSHFYFNLTQANTVDIYRT---NGROY-VLATSTPNYSTPGSAVAYMDIASV 340
QY 446 FNGPFAHRDGPQHOMGCPY-GKVPFPRPGVCPSKMTAQGRPFSTKDYDPDEVLOFARAH 504
Db 341 FTGRFEOKSPDSTWTPVDERVKPRPGCCAGSSILE--RATSNFEDDTLNFITKH 397
QY 505 PLTFMVYRPHRGSPVLYKTLAQOHLQIYVDRVAEDGTVDVIFLGTDSGSLKVALQA 564
Db 398 PLMDEAVPSTFNPFWLRTWRYRLTKIADTAGPYQNTVFLGSEKIIILKFLARIG 457
QY 565 GGSAPPEVYLELOVFK-----VPTIEMEISVRYQMLYGSRLGVNOLRHOC 615
Db 458 NSGFLNDSLFLEEMSYNSKESYDCEVDEKRIQMGLDRASSLYAFSTCVKVLGRC 517
QY 616 ERYGTACACCCALRDYPCAM--DGASCTHYRPSLGRKRRRRODIRGNP-----662
Db 518 ERHGCKTKCTIASRDPYCGMIKEGACSHLSPV--SRLTFEODIERGNMDGLDCHNSFV 575
QY 663 ALCQJGQSGEEBAVGLVATVYGTENSTFLECTPKSPQAA-----VRW--LLQ 710
Db 576 ALNDISTPLPDNE--MSYNTVYG--HSSSLLPSTTSSTQOEGESGMLDKHLDD 630
QY 711 RPDGCP 717
Db 631 SPDSTDP 637

```

RESULT 13
054948 PRELIMINARY: PRT: 284 AA.

AC 054948: 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DB SEMAPHORIN IV HOMOLOG (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hu L.-J., Drabkin H.,
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036162; AAB88861.1;
DR InterPro: IPR001627; Sema.
DR Pfam: PF01403; Sema; 1.
FT NON_TER 284
SQ SEQUENCE 284 AA; 32151 MW; BFD8D7E66BD08027 CRC64;

Query Match 14.4%; Score 685; DB 11; Length 284;
Best Local Similarity 46.2%; Pred. No. 7.8e-49;
Matches 132; Conservative 58; Mismatches 92; Indels 4; Gaps 3;
QY 163 LFTGLDALYSRLDQAMPDPREVLPPOGOREECYRKGRDPLTECANVRYLQPHNRT*222
Db 1 MYGSKDYVLSLDLHDINRPLLIHMAASQRIECLISGKDGNGECGNFVRLIQPNRT*60

```

QY 223 HLLACGTAFOPTCALITYGHG-ENVLHLEPGVSGRGRCPHSPFASTFDGELY 281
Db 61 HLVCYGTANPMCTYVYKRGADYIFLEBEKLESKGCOPYDKLDTASALINELY 120
QY 282 TGLADFLGREAMIFRSGGPPALRSDSOS-LHDPFVMAARIPENDQNDKVFEEF 340
Db 121 AGVYIDFMGTDAIFRTLGKOTAMRTDQNSRWLNDSFIAHLLIPDAENNDKLYEFF 180
QY 341 SEVPSPDGSGNHTVSRGVYCVADAGCQRVLVNKMSTFLKARLVCSVPGGAETHFD 400
Db 181 RER--SAEAPQNPVAYATIGRICLNDGHCCLVNMKSTFLKADLVCSVPGEDGIEETHD 238
QY 401 QLEDFLLMPKAGKSLEYALFSTVASVFGAVCYVHMADIWEV 446
Db 239 ELQDVSVQOTODIRNPVIYAVFTSGSVFRSGAVCYVSMADIRAVF 284

```

RESULT 14

ID 09P249 PRELIMINARY: PRT: 1022 AA.
AC 09P249:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE KIAA1479 PROTEIN (FRAGMENT).
GN KIAA1479.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT Prediction of the coding sequences of unidentified human
RT genes. XVII. The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro.
RL DNA Res. 7:143-150(2000).
DR EMBL: AB040912; BA96003.2;
DR InterPro: IPR003659; PSI.
DR InterPro: IPR001627; Sema.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00423; PSI; 1.
FT NON_TER 1
SQ SEQUENCE 1022 AA; 114372 MW; BEAFBD5FA02C69C4 CRC64;

Query Match 14.4%; Score 682.5; DB 4; Length 1022;
Best Local Similarity 33.2%; Pred. No. 7.5e-48;
Matches 179; Conservative 82; Mismatches 223; Indels 55; Gaps 18;

QY 130 YRDLISANRSAIFLQPSGLNLOAMYLDEYRRLFLGLGIDALYSRLD-----QAMPDR 184
Db 55 FRGRSGNSES-----QHRDQQLML--KIRDTLYAGRDOYVYVNLMEKTEVIVN-K 105
QY 185 EYLPMPQPGQREECYRKGRDPLTECANFVRYLQPHNTHLLACGTAFOPTC---ALITV 241
Db 106 KLTWRSQOODRENCAMKHK-DECHNFIKVFVRPNDEMYVCCGNANPNMCRIYRLSTL 164
QY 242 GHRGEHYHLHPGVSGRGRCPHSPFASTFDGELYGLADFLGREAMIFRSGGP 301
Db 165 EYDGEI-----SLARCPEDAROTNVALPADGKLYSATVADFLASDAVYIRSMGD 215
QY 302 RPAIRS-DSDSLHDPREFVMAARIPENDQNDKVFEEFSETPSPDGSNNHTVSVYG 360
Db 216 GSALRTIKYDSKWKIEPHFLHAI-----EYGVYVFEERE-LAVEHNNIGKAVYSRVA 267
QY 361 RVCYNDAGG-QRVLVNKMSTFLKARLVCSVPGGAETHFDQLEDFLLMPKAGSLEY 419
Db 268 RICNDMGGSQRYLEKHWTSFLKARLVCSV--PGDSFFYFDVLSINDIIOINIP-TVV 324
QY 420 ALFSTVASVFGAVCYVHMADIWEVNGPFAHNDGQHOMGCPY-GKVPFPRPGVCPSK 478

Db 325 GVFITOLNSIPGSAVCAFSMDIDKEVFKREKQKTPDSVTFVAVPEDKVPRKPGCCAKH 384
 QY 479 MTAOGRPGSGKQVDPDVLQOFARAHPLMFVPRPRGRPVLTHTLAQOLHQIVNDRVE 538
 Db 385 GLAE--AKTSIDPDELTSEIKSHPLMSADAVPIADEPFMTKTRVYRLTALISVDHSA 441
 QY 539 AEDGYTVIIFLGTDGSLVKYALIAOAGSAPREVEYLELOVF-----KVPTITE 589
 Db 442 GRYQVTVYFVGSAGMWLVKYLAKTSPFSLN-DSVLEIEIAYNNAKCSAENEDKKYIS 500
 QY 590 MTSYKRWMLVYGSRLGAOLRLHOCCTYGTACAECCCLARDPYCAM-DGASCTHYRPSL 647
 Db 501 LQDLKDHNLVAVFSSCTIRIPLSKCRKSGCKSCIASRDPYCGMLSGSGGRVTPGM 559

RESULT 15

Q91Y36 PRELIMINARY: PRT: 963 AA.
 ID Q91Y36;
 AC Q91Y36;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE SEMAPHORIN 6C.
 GN SEMA6C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C56; TISSUE=BRAIN;
 RA Ou X., Zhai Y., Wei H., Yu Y., Zhang C., Xing G., Zhou K., Zhu Y.,
 RA Ouyang S., He F.;
 RT "Mus musculus mRNA for semaphorin 6C, complete cds."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF363973; AAK52051.1;
 SQ SEQUENCE 963 AA: 102708 MW: 08FA0A3F46532100 CRC64;

Query Match 13.8%; Score 657; DB 11; Length 963;
 Best Local Similarity 29.1%; Pred. No. 9.2e-46;
 Matches 231; Conservative 89; Mismatches 295; Indels 178; Gaps 33;

QY 95 APSAMATCWLGGILLHGGSSGSPSPVPR--LRISYRDLISANRSATFLPGGSLINQ 152
 Db 4 APHSMPILLLL--LL--SSLPQAQAFQDDPTPLITSDLGASPSWMEGLEDVAVA 58
 QY 153 AMYIDEXR-----DRFLGGLDALYSLRDQAMPD-----PREVL-WPQPGQREBCYRK 201
 Db 59 ELGLDPQRFLINRTLLVAARDHVFSDL-QAQEGEGLVNKFILWRSQ--DMENCAYR 115
 QY 202 GRDPLT-ECANFVAVLQPHNRTHLLAGTGAFCPTC--ALITYGHRGHEVHLHEPGSVE 257
 Db 116 GK--LTDECYNYIVLVPMWSQTLACGINSFSPMCRSYGITSLOEGEEL----- 164
 QY 258 SGRGRCPHEPRPASTFIDGELTYGLTADFLGHEAMIFRSGGRPALRSDS-DOSLLHD 316
 Db 165 SGQARCPDPDATQSTVAIFABESLXSATAADQASDAVYRSIGPPLRSKAYDSKWLRE 224
 QY 317 PRFYAARIPENSDDNDKVFSEFVPSIDGGSNHVTVSRGRCVYNDAGO-RVLVN 375
 Db 225 PHEVYAL-----EHGEHYVEFFRE-VVEDARLGRVOFSRVARVCKRDMGSGSPRALDR 276
 QY 376 KWSFTLKARLYCSVPGCGAETHFDQLEDVFLMPKAGKSLEYVALFSTVASVFOGFAVC 435
 Db 277 HMTSLKRLKNCV--PEDSTFEDVLOSITGPNVNLHGRS-ALFGVFTTQINSIPGSAVC 333
 QY 436 VYHMADIWEVFNPFANDGQHOHNGPYG-GKVPFPRPGVCPSKMTAQGRPGSTKDYF 494
 Db 334 AYLDLDIEGFGKFKEDRSIDGAWTPVSEDKVPSRPGSCAGVAA--ASFSSODLP 390
 QY 495 DEVLOFARAHPLMFVPRPRGRPVLTHTLAQOLHQIVNDRVEAEDGTVDYIIFLGTDG 554
 Db 391 DVLLEIFKAHPLLDPAVVPATHQPLITLTSRA-LLTQAVAVDGMAGPHRNTTVLFLGSDNG 449

QY 555 SVLKVIALOAGSAPREVEYLELOVF-----KVPTITEISYKRWMLVYGS 603
 Db 450 TVLKVLIP-PEGOSLGSPPVLEIDAYSHARCSGKSPRARARITIELDTEGHRLFVAF 508
 QY 604 RLGAOLRLHOCCTYGTACAECCCLAR-DPYCAM----- 635
 Db 509 PGCLIVLYLSRCARHG-ACQRCCLASLDPYCGWHRSGCMSIRPGGTVDLTGNQESTE 567
 QY 636 -----DGASCTHYRPS-----LGRFRFR-----QDIRGNPALQCGOSOE----- 672
 Db 568 HGDCQDATGSQSGPDSAYVLLGPGRSPTPSSPSDAQG-PQSTLGAHNGVARDLS 626
 QY 673 -----EAVGLVAATMVYGTENHSTFLDCLPKSPQAAVRLIQRPDGEQDVQKT 722
 Db 627 PASASRSIPILPILACVAAAFALGASVSGLLVSCACRRANRRRSKDIETPG----- 677
 QY 678 -----LPRPLSLRLARLHNG----- 699
 Db 783 EEPFARGGLASTP 795
 Db 700 --PRKDGDAQTP 710

Search completed: October 9, 2002, 14:24:10
 Job time : 75 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 9, 2002, 19:19:59 ; Search time 1640 Seconds
(without alignments)
7201.119 Million cell updates/sec

Title: US-09-813-290-2
Perfect score: 4746
Sequence: 1 MACLAGKVPMPGSMVPMHK.....KMSRVHAENRRPREVEAT 875

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 13736207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV=xlh
-O/cpnt2.1/USPTO.spool/US09813290/rnat_09102002_094646_19135/app_query.fasta.1.1031
-DB-EST -QPM-fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0
-UNITS-bits START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=pic -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09813290_@CGN_1.1.1180_@rnat_09102002_094646_19135 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1625.5	34.2	3436	11	AK014333	Mus muscu
2	820	17.3	647	10	BI067436	pf1n.pko

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE
AK014333	AK014333	Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length cDNA library, clone:3230402E04, short basic domain, secreted, (semaphorin) 3B, full insert sequence.	AK014333	AK014333.1	GI:12852112	HTC: CAP trapper.	Mus musculus (strain: C57BL/6J) 14, 17 days embryo head cDNA to mRNA, clone: libRIKEN full-length enriched mouse cDNA library clone:3230402E04.	Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (sites)	Garnicki, P., and Hayashizaki, Y.	High efficiency full-length cDNA cloning	99279253	2 (sites)	Garnicki, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	Normalization and subtraction of cap-trapper-selected cDNAs to
45	503.5	10.6	701	10	BB615357											

ALIGNMENTS

AK014333 3436 bp mRNA linear HTC 19-JAN-2002
Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
enriched library, clone:3230402E04; sema domain, immunoglobulin
domain (Ig), short basic domain, secreted, (semaphorin) 3B, full
insert sequence.

/db_xref="taxon:9031"
 /clone_id="pfafn.pk012.d10"
 /clone_lib="normalized chicken fat cDNA library"
 /sex="Male and Female"
 /tissue_type="fat"
 /lab_host="E.coli EMDH10B"
 /note="Vector: pSPORI1"

BASE COUNT 119 a 198 c 195 g 123 t 12 others
 ORIGIN

Alignment Scores:

Pred. No.: 3.39e-50 Length: 647
 Score: 820.00 Matches: 157
 Percent Similarity: 82.08% Conservative: 17
 Best Local Similarity: 74.06% Mismatches: 37
 Query Match: 17.28% Indels: 1
 Gaps: 0

US-09-813-290-2 (1-875) x B1067436 (1-647)

QY 353 HisValThrValSerArgValGlyArgValCysValAsnAspAlaGlyGlnArgVal 372
 DB 11 CACACCATCGTCACGCGGTGCAGGGGTCTCGTGAATATGCTGAGAGCCNNAGGGTG 70
 QY 373 LeuValAsnLysTrpSerThrPheLeuLysAlaArgLeuValCysSerValProGlyPro 392
 DB 71 CTGGTCAACAGTGGAGCCCTTCAACAAAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 130
 QY 393 GlyValAlaGlnThrHisPheAspGlnLeuGlnAspValPheLeuLeuTrpProLysAla 412
 DB 131 GCGCGCATGTGATACCATTTTGATGAGCTGAGAGATGCTTTTGGTGGAGACCAAGAT 190
 QY 413 GlyLysSerLeuValValTyrAlaLeuPheSerThrValSerAlaValPheGlnGlyPhe 432
 DB 191 GGAAGAGACCAAGAGATCATGACACTTTCAGCAGCAGTACGATGCTTTTCAAGGCTCC 250
 QY 433 AlaValCysValTyrHisMetAlaAspIleTrpGlnValPheAsnLysProPheAlaHis 452
 DB 251 GCGGTGTGTGTACCGCATGCGGCACATCCGGAAGTCTTCAACGGCCCTTGGCCAC 310
 QY 453 ArgAspGlyProGlnHisGlnTrpGlyProGlyGlyLysValProPheProArgPro 472
 DB 311 CCGGACAGCCCTCCACAGTGGGAGCGCTACGAGAGCGCTGCGCTGCGCTGCGCTGCG 370
 QY 473 GlyValCysProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLysAsp 492
 DB 371 GCGGTGTGTGTACCGCATGCGGCACATCCGGAAGTCTTCAACGGCCCTTGGCCAC 430
 QY 493 TyrProAspGlyValLeuGlnPheAlaArgAlaHisProLeuMetPhe-TrpProValAr 512
 DB 431 CTGGCTGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490
 QY 512 GProArgHisGlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnI 532
 DB 491 CCCAGGAGATACCGCCCGCTGCTGTAAGAGCTGACTGAGCAGCAGCCCTCCACCT 550
 QY 532 eValValAspArgValGlnAlaGlnAspGlyThrTyrAspValIlePheLeuGlyThrAs 552
 DB 551 TGTGTGTGACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 610
 QY 552 pSerGlySerValLeuLysValIleAlaLeuGln 563
 DB 611 TGCAGGCTCGGTGCTGANNNTGTGTGCTGACG 644

RESULT 3
 AM803853/c 420 bp mRNA linear EST 16-MAY-2000
 LOCUS IL2-UM0082-240300-058-F08 UM0082 Homo sapiens cDNA, mRNA sequence.
 DEFINITION AM803853
 ACCESSION AM803853
 VERSION AM803853.1 GI:7855736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

Mammalia: Eutheria: Chordata: Craniata: Vertebrata: Euteleostomi;
 Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
 1 (bases 1 to 420)

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brito, M. R.,
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
 Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J., and
 Simpson, A. J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the RAPSP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=4&t=2-IL2-UM0082-240>)
 Seq primer: puc 18 forward
 High quality sequence start: 3
 High quality sequence stop: 389.
 Location/Qualifiers

FEATURES

source

1..420
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="UM0082"
 /dev_stage="Adult"

/note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESSES PCR (U.S. Letters Patent application No. 196
 ,716 - Ludwig Institute for Cancer Research) profiles
 into the puc 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 86 a 140 c 124 g 70 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.24e-47 Length: 420
 Score: 778.00 Matches: 140
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 16.39% Indels: 0
 Gaps: 0

US-09-813-290-2 (1-875) x AM803853 (1-420)

QY 362 ValCysValAsnAspAlaGlyGlnArgValLeuValAsnLysTrpSerThrPheLeu 381
 DB 420 GTGTGCTGATGATGCTGTGGGCGCAGCGGCTGCTGTGATACAAATGAGCACTTCTC 361
 QY 382 LysAlaArgLeuValCysSerValProGlyProGlyGlyAlaGlnThrHisPheAspGln 401
 DB 360 AAGGCCAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 301
 QY 402 LeuGlnAspValPheLeuLeuTrpProLysAlaGlyLysSerLeuGlnValTyrAlaLeu 421
 DB 300 CTAGAGAGATGTGTTCTGTGTGCTGCTAAGCCCGGAGAGAGCTGAGGTGACCGCTG 241
 QY 422 PheSerThrValSerAlaValPheGlnGlyPheAlaValCysValTyrHisMetAlaAsp 441
 DB 240 TTCAGCAGCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 181
 QY 442 IleTrpGlnValPheAsnLysProPheAlaHisArgAspGlyProGlnHisGlnTrpGly 461
 DB 180 ATCTGGAGAGTTTCAACGGGCCCTTTGCCCCACGAGATGGGCCCTCACACACAGTGGGGG 121

QY 462 ProTyrGlyGlyValPProPheProArgProGlyValAlcysProserLysMetThrAla 481
 DB 120 CCTATGCGGGGAGAGGCGCTTCCCTCCGCGCTGTCGCCACCAAGATGACCGCA 61
 QY 482 G1PProGlyArgProPheGlySerThrLysAspTyrProAspGluValLleuGlnPheAla 501
 DB 60 CAGCCAGAGCGGCTTTGGCAGCAGCAGACAGACTACCATGATGATGTCGCACTTGGC 1
 RESULT 4
 BI652258 990 bp mRNA linear EST 12-SEP-2001
 LOCUS 603299702F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5340217 5'
 DEFINITION mRNA sequence.
 ACCESSION BI652258
 VERSION BI652258.1 GI:15566494
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 990)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHM11864 row: C column: 02
 High quality sequence stop: 892.
 location/Qualifiers
 1. 990
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="5340217"
 /clone_id="NIH_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert 2 kb. Library constructed by Life
 Technologies, catalog #12017-018. Investigators providing
 samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference
 for transgenic model: Xu et al., Nature Genetics 22, 37-43
 (1999). Note: this is a NCI CGAP Library."
 BASE COUNT 194 a 285 c 292 g 219 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.45e-47 Length: 990
 Score: 777.50 Matches: 163
 Percent Similarity: 65.36% Conservative: 54
 Best Local Similarity: 49.10% Mismatches: 95
 Query Match: 16.38% Indels: 21
 DB: 10 Gaps: 6
 US-09-813-290-2 (1-875) x BI652258 (1-990)
 QY 184 ArgGluValLeuTyrProProGlnPro--GlyGln-ArgGluGluCysValArgLysG1 202
 DB 40 AGACGGCTCTGCTACCGCTGAGCGCCCGCTGGAATGGCGAATGCACTGGCGCAGG 99
 QY 202 YArgAspProLeuThrGluCysAlaAsnPheValArgValLeuGlnProHisAsnArgTh 222
 DB 100 GAAGGACATTTGGTACCGAGTGCATGAACCTTCGTAAGAGTGTGCGACGCGCTACAAACACAGC 199
 QY 222 rHisLeuValAlaCysGlyThrGlyAlaPheGlnProThrCysAlaLeuLeuThrValG1 242

DB 160 CCACTTCGTCGCGCTGTGGCAGACGAGCGCTTCCACCAACGCGCGCTTGTGGAGGTGG 219
 QY 242 yHisArg---GlyLunitValLleuHisLeuGlnProGlySerValG1userGlyArgG1 261
 DB 220 CCACCGGCTGGAGAGACCCATGCTTCAACTGACCGGAGAACTTGGAGCGGCAAGG 279
 QY 261 YArgCysProHisGlnProSerArgProPheAlaSerThrPheLleAspGlyGluLeuTy 281
 DB 280 GAACACTCTTATACCAACCAAGCGATCGGGCTCGGTGCTGGGGGAAGAGCTCTA 339
 QY 281 rThrGlyLeuThrAlaAspPheLeuGlyArgGluAlaMetLlePheArgSerGlyGlyPr 301
 DB 340 TTCTGGGGTGGCAGCAGACACTTATGAGGGCGGAGCTTACCAATCTTTCACACCTTGGCA 399
 QY 301 oArgProAlaLeuArgSerAsp---SerAspGlnSerLeuLeuHisAspProArgPheVa 320
 DB 400 GAATCCGAGTCTCCGAMACAGAGCCCAATGATTCGCCGCTGCTCAATGACCAAGTTGT 459
 QY 320 lMetAlaAlaArgLleProGluAsnSerAspGlnAspAsnAspLysValTyrPhePheP 340
 DB 460 CAAGTCTTTGGATTCACAGAGAGTAGACCCCTGATGACGATTAATCTATTCTTCTT 519
 QY 340 eSerGluThrValProSerProAspGlyGlySerAsnHisValThrValSerArgValG1 360
 DB 520 CCGGAGTCCCGCTGTGGAAGCAGCAGCAGATGGGGCGCATGTGTGTCTGTGTGG 579
 QY 360 YArgValCysValAsnAspAlaGlyGlyArgValLeuValAsnLysTyrPserThrPh 380
 DB 580 CCAGATCTGCGAGAGATGACCTGGGTGGCGGAGCTTGTGTGCAACAAATGACACACTT 639
 QY 380 eLeuLysAlaArgLeuValCysSerValProGlyProGlyGlyAlaGluThrHisPheAs 400
 DB 640 TCTGAAGCGCGGCTGTGTGTCTCAGTCTGAGTGTAGGAGT---GACACCCACTTGTGA 696
 QY 400 pGlnLeuGluAspValPheLeuLeuTyrProLysAlaGlyLysSerLeuGluValTyrAl 420
 DB 697 CCACCTTGAGAGVGTTCCTCTCTCTCTCTCCGAGACCGCCAGACCTCTCTATGTC 756
 QY 420 aLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCysValTyrHisMet-A 440
 DB 757 TGTCTTCTCCACCTCCAGTGTCTCTCCAGGCGCTGTGTGTGTGTATACAGATGA 816
 QY 440 lAspPleTPrGluValPheAsnGlyProPheAlaHisArgAspGlyProGlnHisGlnT 460
 DB 817 CCGATGTGGCGGAGACTTCTGGGAGCCTATAGGCTCAACAGAGGCGCTTACACACACT 876
 QY 460 rGlyProGlyArgLysValPProPheProArgProGlyValCysProserLysMet 480
 DB 877 GGGTGTCTACCGGGGTGGGTCTTATCCAGACTGATGTGCCCCAGACAAACTT 936
 QY 480 hValGlnProGlyArgProPheGlySerThrLysAspTyrProAspGluValLeuGlnP 500
 DB 937 GGGCC-----CTTGGCTCCACAGAGGC-TTCCGAATGGGCTTCCCGCT- 979
 QY 500 heAlaArgAlaHisProLeuMetPheTyrPro 510
 DB 980 -----TGGCGG 985
 RESULT 5
 BI689389 785 bp mRNA linear EST 18-SEP-2001
 LOCUS 60315834F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:535609 5'
 DEFINITION mRNA sequence.
 ACCESSION BI689389
 VERSION BI689389.1 GI:15652018
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 785)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9qabs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM1904 row: d column: 10
High quality sequence stop: 783.

FEATURES
source Location/Qualifiers

1..785
/organism="Mus musculus"
/strain="PVB/N"
/db_xref="taxon:10090"
/clone_image="5355609"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies, Inc. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 161 a 239 c 200 g 185 t

ORIGIN

Alignment Scores:

Pred. No.: 6.09e-45 Length: 785
Score: 749.50 Matches: 152
Percent Similarity: 71.59% Conservative: 37
Best Local Similarity: 57.58% Mismatches: 68
Query Match: 15.79% Indels: 7
DB: 10 Gaps: 2

US-09-813-290-2 (1-875) x B1689389 (1-785)

354 ValThrValSerArgValGlyArgValCys-ValAsnAspAlaGlyGlyGlnArgValLe 373
Db 5 ATGCTGTGTCTCTGTTGGCCACATCTGCGAGAGAGTACCTGGGTGGCGAGGCTT 64
373 UValAsnLysTrpSerThrPheLeuLysAlaArgLeuValCysSerValProGlyProG 393
Db 65 GGTCAACAAATGAGACCACTTCTGAAGGCCGCGCTGTGTCTCAGTACCTGGAGTTGA 124
393 YGlyAlaGluThrHisPheAspGlnLeuGluAspValPheLeuLeuTrpPolySAlaG 413
Db 125 GGCT--GACACCCACTTGCACCACTTCAGATGTTTCTTCTTCCCTCCGAGACCG 181
413 YLysSerLeuGluValTyraLysAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAl 433
Db 182 CCAAGACACCTTCTCTATGCTGTCTTCCACCTCCAGTGTCTTCCAGGCTCTGC 241
433 aValCysValTyHisMetLysAspLeuTrpGluValPheAsnGlyProPheAlaHisAr 453
Db 242 TGTGTGTGTCTACAGCATGTGACGATGTGCGCCGAGCTTCTTGGACCTTTTGTCTCA 301
453 GAAPGlyProGlnHisGlnTrpGlyProTyrglyGlyLysValProPheProAlaProG 473
Db 302 AGAGGGGCTTACACACAGTGGGTCTTCCAGAGGTGTCCTTCCCAACCAAGACCTGG 361
473 yValCysProSerLysMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTy 493
Db 362 CATGTGCCCAAGACAGCTTGGC-----ACCTTACGCTCCACCAAGAGCTT 409
493 rProAspGluValLeuGlnPheAlaArgAlaHisProLeuMetPheTrpProValaLarPr 513
Db 410 CCCAGATGAGCTTATCCAGTTTGTCTGGAACACACCTCTCATGTACAAACCAAGTCTGCC 469

513 oArGHisGlyArgProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleVa 533
Db 470 CATGGGGGGGCGCCCTCTCTCTACAACTGGAGCTGGGTACACCTTCACCCAAATCCG 529
533 lValAspArgValGluAlaGluAspGlyThrTyraAspValIlePheLeuGlyThrAsp 553
Db 530 CCAAGACCAAGTACAGACGCTGGCGATGACACATGATGTCTTCTTATTTGATGATGT 589
553 rGlySerValLeuLysValIleAlaLeuGlnAlaGlyGlySerAlaGluProGlnGluVa 573
Db 590 GGGCACAGTGTGAAAGTGTCTCAGTCCCAAGGACCGACCTAATTCATGAAGGACT 649
573 lValLeuGluGluLeuGlnValPheLysValProThrProIleThrGlnMetGlnIleSe 593
Db 650 TCTCTCGAAGAGCTGACAGTGTGTGAGGACTCTGCGCTATCACCAAGCATCAATCTC 709
593 rValLysArgGlnMetLeuTyraLysValGlySerArgGlyValAlaGlnLeuAspGluH 613
Db 710 CTCTAAAGCAACAACCTTACGTAAAGCATGCCGACCGACGATGGCCGATGCTTTC 769
613 lSgLnCys 615
Db 770 ATCGCTGC 777

RESULT 6 645 bp mRNA linear EST 15-JUN-2001
B1066459
LOCUS B1066459

DEFINITION pGln.pk008.d21 normalized chicken fat cDNA library Gallus gallus
CDNA clone pGln.pk008.d21 5' similar to g119910362
reflmp.064548.11 semaphorin sem2 [Homo sapiens] g111449696
reflmp.003233.11 semaphorin sem2 [Homo sapiens] dbj|BA8132.11
(AB029496) semaphorin sem2 [Homo sapiens]g, mRNA sequence.
B1066459
B1066459.1 GI:14473981

VERSION B1066459.1
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 645)
Cogburn,L.A., Morgan,R.W. and Burnside,J.
Chicken ESTs from fat
Unpublished (2001)
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES
source Location/Qualifiers
1..645
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_image="pGln.pk008.d21"
/clone_lib="normalized chicken fat cDNA library"
/sex="Male and Female"
/tissue_type="fat"
/lab_host="E.coli EMDH10B"
/note="Vector: pSPORT1"

BASE COUNT 98 a 221 c 202 g 118 t 6 others

ORIGIN

Alignment Scores:
Pred. No.: 3.31e-42 Length: 645
Score: 710.50 Matches: 137
Percent Similarity: 74.30% Conservative: 22
Best Local Similarity: 64.02% Mismatches: 44
Query Match: 14.97% Indels: 11
DB: 10 Gaps: 4

US-09-813-290-2 (1-875) x B1066459 (1-645)

OY 98 AATP---AlaIleCysThrLeuLeuGlyLeuLeuHisGlyIleSerGly 116
 DB 6 GCGTGGGACGCGCTGGCTGGGGGCGCGCTG-----GCCCGGGG 53
 OY 117 ProSerProGlyProSerValProArgLeuArgLeuSerTyrArgAspLeuSerAla 136
 DB 54 AGCAGCCCG-----CCCGGCTCCCGCTCCCTACCGGAGCTCTGGGTATG 101
 OY 137 AsnArgSerAlaIlePheLeuGlyProGlnGlySerLeuAsnLeuGlnAlaMetTyrLeu 156
 DB 102 AACCCCTCGCTCTTCTTCTTGGCCAGGGGCTTCTGGCTCCGCTCCCTTACCTG 161
 OY 157 AspGlyTyrArgAspArgLeuPheLeuGlyGlyLeuAspAlaLeuTyrSerLeuArgLeu 176
 DB 162 GATGAGTACCGGAGCGCTCTTCACTGGGGGCAAGAGTCTCTACTCCCTCGCTG 221
 OY 177 AspGlnAlaTyrProAspProArgLeuValLeuTyrProProGlnProGlyIleArgIle 196
 DB 222 GACAGGCGCAGCGCGGACACCAAGAGATCTACTGCGCCGCTCTGGAGACAGCGAG 281
 OY 197 GluGlyValArgGlyArgAspProLeuThrGluCysAlaAsnPheValArgValLeu 216
 DB 282 GAGTGTCTTGGAGAGGAGAGACCCAGACCGAGCTGTGCACTAGCTCCGCTGCTG 341
 OY 217 GlnProHisAsnArgThrHisLeuLeuAlaCysGlyThrGlyAlaPheGlnProThrCys 236
 DB 342 CAGCCCTACACAGAACACACCTGCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 401
 OY 237 AlaLeuIleThrValGlyHisArgGlyIle-----HisValLeuHisLeuGlnProGly 254
 DB 402 ACCCTCATCTAGCTGGGAGCGCGGCTGAGACCCAGACCTTCCAGCTTGGAGCCTGCTG 461
 OY 255 SerValGlnSerGlyArgGlyArgCysProHisGlnProSerArgProPheAlaSerThr 274
 DB 462 AGCAGCGAAGAGT 521
 OY 275 PheIleAspGlyLeuLeuTyrThrGlyLeuThrAlaAspPheLeuGlyArgGlnAlaMet 294
 DB 522 ATCATGCGGAGGAGCTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 581
 OY 295 IlePheArgSerGlyIleProArgProAlaLeuArgSerAsp 308
 DB 582 ATCTTCCGAGACCGGAGCGCGCGCTGCCTGCCTGCCTGCCTGCCTGCCTGCCTGC 623
 RESULT 7
 LOCUS B1557438 662 bp mRNA linear EST 05-SEP-2001
 DEFINITION 60338962F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5292039 5',
 mRNA sequence.
 ACCESSION B1557438
 VERSION B1557438.1 GI:15444752
 KEYWORDS house mouse.
 SOURCE EST.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (Bases 1 to 662)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1AM1738 row: k column: 16
 High quality sequence stop: 660.
 Location/Qualifiers
 1..662
 FEATURES
 source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5292039"
 /clone_lib="NIH_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert 2 kb. Library constructed by Life
 Technologies, catalog #12017-018. Investigators providing
 samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference
 for transgenic model: Xu et al., Nature Genetics 22, 37-43
 (1999). Note: this is a NCI_CGAP Library."
 BASE COUNT 134 a 209 c 168 g 151 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,69e-41 Length: 662
 Score: 701.00 Matches: 131
 Percent Similarity: 70.98% Conserved: 28
 Best Local Similarity: 58.48% Mismatches: 61
 Query Match: 14.77% Indels: 4
 DB: 10 Gaps: 1
 US-09-813-290-2 (1-875) x B1557438 (1-662)
 OY 418 ValTyrAlaLeuPheSerThrValSerAlaValPheGlnGlyPheAlaValCysValTyr 437
 DB 2 CTCTATGCTGTCTTCTCCACCTCCAGTGGTGTCTTCCAGGCGCTGTGTGCTGCTAC 61
 OY 438 HisMetAlaAspIleThrPrgIleValPheAsnGlyProPheAlaHisArgAspGlyProGln 457
 DB 62 AGCATGACCATGTGGCGCGAGGCTTCTTGGGACCTTCTCACAAGAGGGGCGCTACA 121
 OY 458 HisGlnTyrProTyrGlyGlyValPheProPheProArgProGlyValCysProSer 477
 DB 122 CACCAGTGGGTGTCTTACCAAGGCTGTCTTCCCTACCAAGAGCTGGCATGTGCCAGC 181
 OY 478 LysMetThrAlaGlnProGlyArgProPheGlySerThrLysAspTyrProAspGlyVal 497
 DB 182 AAGACCTTTGGC-----ACCTTCAGCTCCCAAGAGCATTCACAGATGACGTT 229
 OY 498 LeuGlnPheAlaArgAlaHisProLeuMetPheThrProValArgProArgHisGlyArg 517
 DB 230 ATCCAGTTTCTCGGAACCACTCTCATGTACCAACCAAGCTGCCCATGGGGGCGC 289
 OY 518 ProValLeuValLysThrHisLeuAlaGlnGlnLeuHisGlnIleValAlaAspArgVal 537
 DB 290 CTTCTTCTTCTACCAAGTGGGAGCTGGGTACACCTTACCCCAATCGCCGACGAGTA 349
 OY 538 GluAlaGlnAspGlyThrTyrAspValIlePheLeuGlyThrAspSerGlySerValLeu 557
 DB 350 GCAGCTCCGATGACACTGATGTCTTTCATTTGTATGATGATGTGGGACAGTCTG 409
 OY 558 LysValIleAlaLeuGlnAlaGlyGlySerAlaGlnProGlnGlnValValLeuGln 577
 DB 410 AAAGTGTATGATCCCAAGAGGAGCGAGCTTATTTCTGAAGACTTCTCTGAAGAG 469
 OY 578 LeuGlnValPheLysValProThrProIleThrGlnMetGlnLeuSerValLysArgGln 597
 DB 470 CTGAGAGGTTCGAGGAGCTGCGCGCTATACACAGATGAATCTCTTAAGAGCAA 529
 OY 598 MetLeuTyrValGlySerArgLeuGlyValAlaGlnLeuArgLeuHisGlnCysGlnThr 617
 DB 530 CAACCTCAGTACGATGATGCGGAGCGAGTGGCCAGATGCTTTGATCAGCTGAGCC 589
 OY 618 TyrGlyThrAlaCysAlaGlnCysCysLeuAlaArgAspProTyrCysAlaTyrAspGly 637
 DB 590 CTAGGCGCTGCTGGCAGAAATCTGCTTGGCCCGCTGATCTTACTGCGCTGGGATGA 649
 OY 638 AlaSerCysThr 641
 DB 650 TCAGTTTCACA 661

RESULT 8
AL584625
LOCUS
DEFINITION AL584625 Stratagene Chick Embryo Lambda cDNA Library (* 937405)
ACCESSION AL584625
VERSION AL584625.1 GI:13163356
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 754)
REFERENCE
AUTHORS Murray, F.
TITLE Stratagene Chick Embryo Lambda cDNA Library
JOURNAL Unpublished (2001)
COMMENT Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
Seq primer: T3.
location/Qualifiers
1. 754
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_lib="Ros012B12"
/clone_lib="Stratagene Chick Embryo Lambda cDNA Library (* 937405)"
/tissue_type="Embryo"
/dev_stage="5 days old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="vector: pBLUESCRIPT SK⁺ Site-1: EcoRI; Site-2: XhoI
; Cloned unidirectionally. Primer: Oligo dT, Uni-ZAP XR
vector. Average insert size: 1.5kb.; 5' adaptor sequence:
5' GAATTCGGCAGCAG 3'; 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 158 a 242 c 218 g 134 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 8,366-41 Length: 754
Score: 692.50 Matches: 138
Percent Similarity: 68.80% Conservative: 34
Best Local Similarity: 55.20% Mismatches: 66
Query Match: 14,598 Indels: 13
DB: 9 Gaps: 5

US-09-813-290-2 (1-875) x AL584625 (1-754)

QY 423 SerThValSerAlaValPheGlnGlyPheAlaValCysValTyrHisMetAlaAspIle 442
DB 15 TCCGCGATCGGGGTGGGTGTTCAAGGCTCGGCTGCTGCTTACCTCCAGCGCGACATC 74
QY 443 TrpGluValPheAsnGlyProPheAlaHisArgAspGlyProGlnHisGlnTrpGlyPro 462
DB 75 CCGATGCTGTCAATGAGGCGCTCGCGCACAAAGAGGCGCCCAACACACAGTGGATGCC 134
QY 463 TyrGlyGlyValPheProPheProAlaTrpGlyValCysProSerLysMetThrAlaGln 482
DB 135 TACAGGGGAAATACCAACCCCGCGCGACCTGCGCGGGGG--ACCTTCACA 191
QY 483 ProGlyArgProPheGlySerThrLysAspTyrProAspGlnValLeuGlnPheAlaArg 502
DB 192 CCCTCC-----ATGAAGTCAACCAAGAGACTACCCGACAGAGTGAATCAACTTCATGCCG 245
QY 503 AlaHisProLeuMetPheThrProValArgProArgHisGlyArgProValLeuValLys 522
DB 246 TCCACACCTCTCATGTACACACCGCGCTACCGCGCCACCGGACGCTCTGCTGCTGCCG 305

QY 523 ThrHisLeuAlaGlnGluLeuHisGlnIleValAlaAspArgValGluAlaGlyAspGly 542
DB 306 ACCAAGCTCACTACCGCTTCACACACATCGCTGTGACAGGTGAGCGGACGATCGG 365
QY 543 ThrTyrAspValIlePheLeuGlyThrAspSerGlySerValLeuLysValIleAlaLeu 562
DB 366 CGCTATGAGTGTCTTCTTCTGCGCAGATCGGGGACAGTGCAGAAAGTATTGTGCTG 425
QY 563 GlnAlaGlyGlySerAlaGluProGluGluValValLeuGluGluLeuGlnValPheLys 582
DB 426 CCCCCG---GATGACATGTAGACAGACAGAGAGCTAGTGCAGAGATGAGTGTCTAG 482
QY 583 ValProThrProIleThrIleGluMetGluLysSerValLysArgGlnMetLeuTyrValGly 602
DB 483 GTGCCAGCACCCTATCAAGATGATGACCATCTCTCCACAGAGCAACACTTACGTGCC 542
QY 603 SerArgLeuGlyValAlaGlnLeuArgPheHisGlnCysGluThrTyrGlyThrAlaCys 622
DB 543 TCGGCAGTAGAGAGTACCCACCTGCGCTGCACCGGTGTGACGTGTATGAGAAAGCTGT 602
QY 623 AlaGluCysCysLeuAlaArgAspProTyrCysAlaIleTrpAspGlyAlaSerCysThrHis 642
DB 603 GCTACTGCTGCT--GCCCGGACCCATACTGTGCTGTGGATGGCAGGCGCTGCAGCGCC 660
QY 643 TyrArgProSerLeuGlyLysArgArgPheArgArgGlnAspIleArgHisGlyAsnPro 662
DB 661 TAC-----TCACATCTCTCAAGCGGACAGCAGGAGCAGTCCGACGACGACCC 714
QY 663 -----AlaLeuGlnCys 666
DB 715 ATGCGCGGTGCGGAGTACAACTTCATATGC 744

RESULT 9
BI067451
LOCUS
DEFINITION BI067451 576 bp mRNA linear EST 15-JUN-2001
cDNA clone pgfin.pk012.d9 normalized chicken fat cDNA library gallus
collapsin-5 [gallus gallus]g, mRNA sequence.
ACCESSION BI067451
VERSION BI067451.1 GI:14474973
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 576)
REFERENCE
AUTHORS Cogburn, L.A., Morgan, R.W. and Burnside, J.
TITLE Chicken ESTs from fat
JOURNAL Unpublished (2001)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
location/Qualifiers
1. 576
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_lib="pgfin.pk012.d9"
/clone_lib="normalized chicken fat cDNA library"
/sex="Male and Female"
/tissue_type="fat"
/lab_host="E.coli EMDH10B"
/note="vector: pSPORT1"

BASE COUNT 106 a 179 c 161 g 104 t 26 others
ORIGIN

Alignment Scores:
Pred. No.: 8,966-41 Length: 576
Score: 690.00 Matches: 135

[illegible]

